# 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: December 2011

*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 censoring, a nonparametric conditional quantile with censoring, an interval-outcome partially-linear regression, an interval-outcome nonparametric regression, a semiparametric discrete-choice model with multiple equilibria, a nonparametric revealed preference model, and tests of a variety of functional inequalities, including nonparametric average treatment effects for certain sub-populations.

As far as we are aware, the only other paper in the literature that covers the examples described above is Chernozhukov, Lee, and Rosen (2008) (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.

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 (2007a,b) (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 censoring 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 Jia (2008), 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.

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 (2010b).

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 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 (2008), Kim (2008), Aradillas-López, Gandhi, and Quint (2010), Beresteanu, Molchanov, and Molinari (2010), Ponomareva (2010), Armstrong (2011a,b), Hsu (2011), and Lee, Song, and Whang (2011). 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

[1] This holds because the functions $g_1(x), ..., g_k(x)$ in (3.1) below, which multiply the moment functions indexed by $1, ..., k$, need not be the same.

moment inequalities of a particular type. Menzel (2008) 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 give pointwise results. The present paper provides asymptotic results under fixed and some local alternatives, whereas CLR do not give results under the alternative.

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 considered in the paper. Section 4 describes the critical values considered with the focus on GMS critical values. Section 5 establishes the uniform asymptotic coverage probabilities of the CS's. Sections 6 and 7 establish the power of the tests against fixed and some local alternatives, respectively. Section 8 provides Monte Carlo simulation results for two models. Appendix 1 provides proofs of the uniform asymptotic size results. For brevity, Appendix 2 is given in Andrews and Shi (2010a). It provides proofs of the results under fixed and some local alternatives and gives 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:

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

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 $\theta_0$. We do not assume that $\theta_0$ 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 $\theta_0$ 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 $\theta$ 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 $\alpha$. Then, a nominal level $1-\alpha$ CS for the true value $\theta_0$ 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 8.

**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. The "parametric" version of this example—where $Z_i$ is not present—is discussed in Chernozhukov, Lee and Rosen (2008). 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 $\tau$ 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 $z_0 \in Supp(Z)$,

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

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 $z_0 \in Supp(Z_1)$:

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

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 (2008), and Andrews and Shi (2007b)). 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, $\varepsilon_j$ 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 $\alpha_0$. Let $F_1$ and $F_2$ denote the marginal distributions of $\varepsilon_1$ and $\varepsilon_2$ 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:

$$\begin{aligned}
E((1-D_1)(1-D_2) - P_{00}(X,\theta_0)|X, Z = z_0) &= 0, \\
E(D_1D_2 - P_{11}(X,\theta_0)|X, Z = z_0) &= 0, \\
E(D_1(1-D_2) - P_{10}(X,\theta_0)|X, Z = z_0) &\geq 0, \text{ and} \\
E(D_2(1-D_1) - P_{01}(X,\theta_0)|X, Z = z_0) &\geq 0, \qquad (2.7)
\end{aligned}$$

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), $\psi_{j0}$ 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 $\psi_{j0}$ 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 guarantees that (2.9) implies a moment inequality model of the following form:

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

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

$$r(a_j^*, a_{-j}, W) - r(a_j, a_{-j}, W) = G(a_j^*, a_j, a_{-j}, \beta_0, X, \psi_0(Z)), \tag{2.11}$$

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 $\theta$ 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 $\theta$) 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:

$$\begin{aligned} H_0 \ : \ u_j(x, z_0) \geq 0 \text{ for } z_0 \in \mathcal{Z} \text{ and all } (x, j) \in \mathcal{X} \times \{1, ..., p\}, \text{ where} \\ u_j(x, z) = E(m_j(W_i)|X_i = x, Z_i = z) \end{aligned} \tag{2.12}$$

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 $p(X_i, Z_i)$, 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 $i$th individual is the difference between $y_i(1)$ and $y_i(0)$. The

[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.

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 (2.12) can be applied with $p = 1$ and

$$m(W_i) = \frac{D_i Y_i}{p(X_i, Z_i)} - \frac{(1 - D_i)Y_i}{1 - p(X_i, Z_i)}, \tag{2.13}$$

where $W_i = (Y_i, D_i, X_i, Z_i)$ and no parameter $\theta$ appears in the problem. $\square$

## 2.3 Parameter Space

Let $(\theta, F)$ denote generic values of the parameter and distribution. Let $\mathcal{F}$ denote the parameter space for $(\theta_0, F_0)$. 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 $(\theta, F)$. Let $F_{X|z}$ denote the conditional distribution of $X_i$ given $Z_i = z$ under $(\theta, F)$. Let $F_Z$ and $F_X$ denote the marginal distributions of $Z_i$ and $X_i$, respectively, under $(\theta, F)$.

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

Define

$$\begin{aligned} m_F(\theta, x, z) &= E_F(m(W_i, \theta)|X_i = x, Z_i = z) f(z|x), \\ \Sigma_F(\theta, x, z) &= E_F(m(W_i, \theta)m(W_i, \theta)'|X_i = x, Z_i = z) f(z|x), \text{ and} \\ \sigma^2_{F,j}(\theta, z) &= E_F(m_j^2(W_i, \theta)|Z_i = z) f(z) \text{ for } j \leq k, \end{aligned} \tag{2.14}$$

where $f(z|x)$ 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 $\mu_{Leb}$ on $\mathcal{Z}_0$, defined in (2.15) below.

The parameter space $\mathcal{F}$ is defined to be the collection of $(\theta, F)$ that satisfy the following conditions:

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

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

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

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

(v) $F_Z$ restricted to $z \in \mathcal{Z}_0$ is absolutely continuous wrt $\mu_{Leb}$ with density $f(z)$ $\forall z \in \mathcal{Z}_0$,

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

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

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

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

(x) $\sigma^2_{F,j}(\theta, z_0) \geq \delta_j$ for $j \leq k$,

and

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

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

(xiii) $\Sigma_F(\theta, x, z)$ 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

(xiv) $E_F\left(|m_j(W_i, \theta)|^4 \middle| Z_i = z\right) f(z) \leq C_4$ $\forall z \in \mathcal{Z}_0$ $\forall j \leq k$, (2.15)

for some $C_\ell < \infty$ for $\ell = 1, ..., 4$ and $\delta, \delta_j > 0$ for $j \leq k$, where $k = p + v$.

Conditions (iii) and (iv) of $\mathcal{F}$ are the key partial-identification conditions of the model. Conditions (v)-(ix) of $\mathcal{F}$ are absolute continuity conditions. Conditions (v) and (viii) 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 $z_0$ the results of AS1 cover the model. Conditions (vi), (vii), and (ix) are not very restrictive because the absolute continuity is wrt arbitrary measures $\mu_X$ and $\mu_Y$. Conditions (x)-(xiv) bound some variances away from zero and impose 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 Test Statistics

## 3.1 Form of the Test Statistic

Next, 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:

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

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

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

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

The collection $\mathcal{G}$ is chosen so that $\Theta_{F_0}(\mathcal{G}) = \Theta_{F_0}$, defined in (2.2). Section 3.3 provides conditions for this equality and gives an example of an instrument set $\mathcal{G}$ that satisfies the conditions. Additional sets $\mathcal{G}$ are given in AS1 and AS2.

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

$$
\begin{aligned}
\overline{m}_n(\theta, g) &= n^{-1} \sum_{i=1}^{n} m(W_i, \theta, g, b) \text{ for } g \in \mathcal{G}, \text{ where} \\
m(W_i, \theta, g, b) &= b^{-d_z/2} K_b(Z_i) m(W_i, \theta, g), \\
K_b(Z_i) &= K\left(\frac{Z_i - z_0}{b}\right), \\
m(W_i, \theta, g) &= \begin{pmatrix} m_1(W_i, \theta) g_1(X_i) \\ m_2(W_i, \theta) g_2(X_i) \\ \vdots \\ m_k(W_i, \theta) g_k(X_i) \end{pmatrix} \text{ for } g \in \mathcal{G},
\end{aligned} \tag{3.3}
$$

$b > 0$ is a scalar bandwidth parameter for which $b \to 0$ as $n \to \infty$, and $K(x)$ is a kernel function. The definition of $\overline{m}_n(\theta, g)$ in (3.3) 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 $m(W_i, \theta, g, b)$.

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.[4]

We assume the bandwidth $b$ and kernel function $K(x)$ satisfy:

**Assumption B.** (a) $b = o(n^{-1/(4+d_z)})$ and (b) $nb^{d_z} \to \infty$ as $n \to \infty$.

**Assumption K.** (a) $\int K(z) dz = 1$, (b) $\int z K(z) dz = 0_{d_z}$, (c) $K(z) = 0 \ \forall z \notin [-1, 1]^{d_z}$, (d) $K(z) \geq 0 \ \forall z \in R^{d_z}$, and (e) $\sup_{z \in R^{d_z}} K(z) < \infty$.

Assumptions B and K are standard assumptions in the nonparametric density and regression literature. When Assumption B is applied to a nonparametric regression or density estimator, part (a) implies that the bias of the estimator goes to zero faster than

[4]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 variance (and is the weakest condition for which this holds) and part (b) implies that the estimator is asymptotically normal (because it implies that $b$ goes to zero sufficiently slowly that a Lindeberg condition holds).

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

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

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

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

for some fixed $\varepsilon > 0$. In the simulations in Section 8, we use $\varepsilon = 5/100$. By design, $\overline{\Sigma}_n(\theta, g)$ is a linear combination of two scale equivariant functions and hence is scale equivariant.[5] This yields a test statistic that is invariant to rescaling of the moment functions $m(W_i, \theta)$, which is an important property.

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

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

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

The Kolmogorov-Smirnov-type (KS) statistic is

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

For brevity, the discussion in this paper focusses on CvM statistics and all results

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

stated concern CvM statistics. Similar results hold for KS statistics. 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 E of AS2.

## 3.2 S Functions

We establish the results of this paper for a broad family of functions $S$. For brevity, the conditions on $S$, viz., Assumptions S1-S4, are stated in Appendix 1. Three leading functions that satisfy these conditions are:

$$
\begin{aligned}
S_1(m,\Sigma) &= \sum_{j=1}^{p}[m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v}[m_j/\sigma_j]^2, \\
S_2(m,\Sigma) &= \inf_{t=(t_1',0_v')':t_1\in R_{+,\infty}^p}(m-t)'\Sigma^{-1}(m-t), \text{ and} \\
S_3(m,\Sigma) &= \max\{[m_1/\sigma_1]_-^2,...,[m_p/\sigma_p]_-^2,(m_{p+1}/\sigma_{p+1})^2,...,(m_{p+v}/\sigma_{p+v})^2\},
\end{aligned} \tag{3.8}
$$

where $m_j$ is the $j$th element of the vector $m$, $\sigma_j^2$ is the $j$th diagonal element of the matrix $\Sigma$, 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.[6] The functions $S_1$, $S_2$, and $S_3$ are the modified method of moments (MMM) or Sum function, the quasi-likelihood ratio (QLR) function, and the Max function, respectively.

## 3.3 X-Instruments

The collection of instruments $\mathcal{G}$ needs to satisfy the following condition in order for the conditional moments $\{E_F(m(W_i,\theta,g)|Z_i = z_0) : g \in \mathcal{G}\}$ to incorporate the same information as the conditional moments $\{E_F(m(W_i,\theta)|X_i = x, Z_i = z_0) : x \in R^{d_x}\}$.

For any $\theta \in \Theta$ and any distribution $F$ with $E_F(||m(W_i,\theta)||\ |Z_i = z_0) < \infty$, let

$$
\begin{aligned}
\mathcal{X}_F(\theta) = \{x \in R^{d_x} : \ & E_F(m_j(W_i,\theta)|X_i = x, Z_i = z_0) < 0 \text{ for some } j \leq p \text{ or} \\
& E_F(m_j(W_i,\theta)|X_i = x, Z_i = z_0) \neq 0 \text{ for some } j = p+1,...,k\}.
\end{aligned} \tag{3.9}
$$

**Assumption NCI.** For any $\theta \in \Theta$ and distribution $F$ for which $E_F(||m(W_i,\theta)||\ |Z_i =$

[6] The functions $S_1$, $S_2$, and $S_3$ satisfy Assumptions S1-S4, stated in Appendix 1, by Lemma 1 of AS1.

$z_0) < \infty$ and $P_F(X_i \in \mathcal{X}_F(\theta)|Z_i = z_0) > 0$, there exists some $g \in \mathcal{G}$ such that

$$E_F(m_j(W_i,\theta)g_j(X_i)|Z_i = z_0) < 0 \text{ for some } j \leq p \text{ or}$$
$$E_F(m_j(W_i,\theta)g_j(X_i)|Z_i = z_0) \neq 0 \text{ for some } j = p+1,...,k.$$

Note that NCI abbreviates "nonparametrically conditionally identified." The following Lemma indicates the importance of Assumption NCI.

**Lemma N1.** *Assumption* NCI *implies that* $\Theta_F(\mathcal{G}) = \Theta_F$ *for all* $F$ *with* $\sup_{\theta\in\Theta} E_F(||m(W_i,\theta)||\ |Z_i = z_0) < \infty$.

Collections $\mathcal{G}$ that satisfy Assumption NCI contain non-negative functions whose supports are cubes, boxes, or other sets whose supports are arbitrarily small.

A collection $\mathcal{G}$ must satisfy a "manageability" condition, viz., Assumption NM, that regulates the complexity of $\mathcal{G}$. This condition ensures that $\{n^{1/2}(\overline{m}_n(\theta,g) - E_{F_n}\overline{m}_n(\theta,g)) : g \in \mathcal{G}\}$ satisfies a functional central limit theorem (FCLT) under drifting sequences of distributions $\{F_n : n \geq 1\}$. The latter is utilized in the proof of the uniform coverage probability results for the CS's. The manageability condition is from Pollard (1990) and is defined and explained in Appendix E of AS2. For brevity, Assumption NM is stated in Appendix 1.

Now we give an example of a collection of functions $\mathcal{G}$ that satisfies Assumptions NCI and NM. AS1 and AS2 give four other collections $\mathcal{G}$ that satisfy Assumptions NCI and NM.

**Example. (Countable Hypercubes).** Suppose $X_i$ is transformed via a one-to-one mapping so that each of its elements lies in $[0,1]$. There is no loss in information in doing so. For example, Sections 9 and 10.3.2 of AS1 and Section 13.2 of Appendix B of AS2 provide examples of how this can be done.

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 is countable:

$$\begin{aligned}\mathcal{G}_{c\text{-}cube} &= \{g(x) : g(x) = 1(x \in C) \cdot 1_k \text{ for } C \in \mathcal{C}_{c\text{-}cube}\}, \text{ where}\\ \mathcal{C}_{c\text{-}cube} &= \left\{ C_{a,r} = \prod_{u=1}^{d_x} ((a_u - 1)/(2r), a_u/(2r)] \in [0,1]^{d_x} : a = (a_1, ..., a_{d_x})' \right.\\ &\qquad \left. a_u \in \{1, 2, ..., 2r\} \text{ for } u = 1, ..., d_x \text{ and } r = r_0, r_0 + 1, ... \right\} \end{aligned} \tag{3.10}$$

for some positive integer $r_0$.[7] The terminology "*c-cube*" abbreviates countable cubes. Note that $C_{a,r}$ is a hypercube in $[0,1]^{d_x}$ with smallest vertex indexed by $a$ and side lengths equal to $(2r)^{-1}$.

The class of countable cubes $\mathcal{G}_{c\text{-}cube}$ leads to a test statistic $T_n(\theta)$ for which the integral over $\mathcal{G}$ reduces to a sum. The set $\mathcal{G}_{c\text{-}cube}$ can be used with continuous and/or discrete regressors.

Lemma 3 of AS1 establishes Assumptions NCI and NM for $\mathcal{G}_{c\text{-}cube}$.[8]

## 3.4 Weight Function Q

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

We now give an example of a weight function $Q$ on $\mathcal{G}_{c\text{-}cube}$.

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

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

[8] Lemma 3 of AS1 and Lemma B2 of AS2 also establish Assumptions NCI and NM of this paper for the collections $\mathcal{G}_{box}$, $\mathcal{G}_{B-spline}$, $\mathcal{G}_{box,dd}$, and $\mathcal{G}_{c/d}$ defined there. The proof is the same as in AS2 for Assumptions CI and M with conditioning on $Z_i = z_0$ added throughout.

probability mass function $\{w(r) : r = r_0, r_0 + 1, ...\}$. This yields the test statistic

$$T_n(\theta) = \sum_{r=r_0}^{\infty} w(r) \sum_{a \in \{1,...,2r\}^{d_x}} (2r)^{-d_x} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\Sigma}_n(\theta, g_{a,r})), \tag{3.11}$$

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

## 3.5 Computation of Sums, Integrals, and Suprema

The test statistic $T_n(\theta)$ given in (3.11) involves an infinite sum. A collection $\mathcal{G}$ with an uncountable number of functions $g$ yields a test statistic $T_n(\theta)$ that is an integral with respect to $Q$. This infinite sum or integral can be approximated by truncation, simulation, or quasi-Monte Carlo (QMC) methods. If $\mathcal{G}$ is countable, let $\{g_1, ..., g_{s_n}\}$ denote the first $s_n$ functions $g$ that appear in the infinite sum that defines $T_n(\theta)$. Alternatively, let $\{g_1, ..., g_{s_n}\}$ be $s_n$ i.i.d. functions drawn from $\mathcal{G}$ according to the distribution $Q$. Or, let $\{g_1, ..., g_{s_n}\}$ be the first $s_n$ terms in a QMC approximation of the integral with respect to (wrt) $Q$. Then, an approximate test statistic obtained by truncation, simulation, or QMC methods is

$$\overline{T}_{n,s_n}(\theta) = \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S(n^{1/2}\overline{m}_n(\theta, g_\ell), \overline{\Sigma}_n(\theta, g_\ell)), \tag{3.12}$$

where $w_{Q,n}(\ell) = Q(\{g_\ell\})$ when an infinite sum is truncated, $w_{Q,n}(\ell) = s_n^{-1}$ when $\{g_1, ..., g_{s_n}\}$ are i.i.d. draws from $\mathcal{G}$ according to $Q$, and $w_{Q,n}(\ell)$ is a suitable weight when a QMC method is used. For example, in (3.11), the outer sum can be truncated at $r_{1,n}$, in which case, $s_n = \sum_{r=r_0}^{r_{1,n}} (2r)^{d_x}$ and $w_{Q,n}(\ell) = w(r)(2r)^{-d_x}$ for $\ell$ such that $g_\ell$ corresponds to $g_{a,r}$ for some $a$.

It can shown that truncation at $s_n$, simulation based on $s_n$ simulation repetitions, or QMC approximation based on $s_n$ terms, where $s_n \to \infty$ as $n \to \infty$, is sufficient to maintain the asymptotic validity of the tests and CS's as well as the asymptotic power results under fixed alternatives and most of the results under $n^{-1/2}$-local alternatives. For brevity we do not do so here. The method of proof is analogous to that used in Section 15.1 of Appendix E of AS2 to prove such results stated in Section 13.1 of Appendix B of AS2 for the tests considered in AS1 and AS2.

The KS form of the test statistic requires the computation of a supremum over $g \in \mathcal{G}$.

For computational ease, this can be replaced by a supremum over $g \in \mathcal{G}_n$, where $\mathcal{G}_n \uparrow \mathcal{G}$ as $n \to \infty$, in the test statistic and in the definition of the critical value (defined below). The same asymptotic results for KS tests hold with $\mathcal{G}_n$ in place of $\mathcal{G}$. For results of this sort for the tests considered in AS1 and AS2, see Section 13.1 of Appendix B of AS2 and Section 15.1 of Appendix E of AS2.

# 4 GMS Confidence Sets

## 4.1 GMS Critical Values

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

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

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

where

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

is a mean zero $R^k$-valued Gaussian process with covariance kernel $h_2(\cdot, \cdot)$ on $\mathcal{G} \times \mathcal{G}$, $h_1(\cdot)$ is a function from $\mathcal{G}$ to $R^p_{[+\infty]} \times \{0\}^v$, and $\varepsilon$ is as in the definition of $\overline{\Sigma}_n(\theta, g)$ in (3.5).[9] The definition of $T(h)$ in (4.1) applies to CvM test statistics. For the KS test statistic, one replaces $\int \ldots \, dQ(g)$ by $\sup_{g \in \mathcal{G}} \ldots$ .

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

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

denote the $1-\alpha$ quantile of $T(h)$. If $h_n = (h_{1,n}, h_2)$ was known, we would use $c_0(h_n, 1-\alpha)$ as the critical value for the test statistic $T_n(\theta)$. However, $h_n$ is not known and $h_{1,n}$ cannot be consistently estimated. In consequence, we replace $h_2$ in $c_0(h_{1,n}, h_2, 1-\alpha)$

[9] The sample paths of $\nu_{h_2}(\cdot)$ are concentrated on the set $U^k_\rho(\mathcal{G})$ of bounded uniformly $\rho$-continuous $R^k$-valued functions on $\mathcal{G}$, where $\rho$ is defined in Appendix A of AS2.

by a uniformly consistent estimator $\widehat{h}_{2,n}(\theta)$ $(= \widehat{h}_{2,n}(\theta, \cdot, \cdot))$ of the covariance kernel $h_2$ and we replace $h_{1,n}$ by a data-dependent GMS function $\varphi_n(\theta)$ $(= \varphi_n(\theta, \cdot))$ on $\mathcal{G}$ that is constructed to be less than or equal to $h_{1,n}(g)$ for all $g \in \mathcal{G}$ with probability that goes to one as $n \to \infty$. Because $S(m, \Sigma)$ is non-increasing in $m_I$ by Assumption S1(b) (see Appendix 1), where $m = (m_I', m_{II}')'$, the latter property yields a test whose asymptotic level is less than or equal to the nominal level $\alpha$. (It is arbitrarily close to $\alpha$ for certain $(\theta, F) \in \mathcal{F}$.) The quantities $\widehat{h}_{2,n}(\theta)$ and $\varphi_n(\theta)$ are defined below.

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

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

where $\eta > 0$ is an arbitrarily small positive constant, e.g., .001. A nominal $1 - \alpha$ GMS CS is given by (2.3) with the critical value $c_{n,1-\alpha}(\theta)$ equal to $c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1 - \alpha)$.[10]

Next, 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[11]

$$\begin{aligned} h_{2,F}(\theta, g, g^*) &= D_F^{-1/2}(\theta, z_0)\Sigma_F(\theta, g, g^*, z_0)D_F^{-1/2}(\theta, z_0), \text{ where} \\ \Sigma_F(\theta, g, g^*, z) &= E_F(m(W_i, \theta, g)m(W_i, \theta, g^*)'|Z_i = z)f(z) \text{ and} \\ D_F(\theta, z) &= Diag(\Sigma_F(\theta, 1_k, 1_k, z)) \ (= Diag(E_F(m(W_i, \theta)m(W_i, \theta)'|Z_i = z)f(z))). \end{aligned} \tag{4.5}$$

Correspondingly, the sample covariance kernel $\widehat{h}_{2,n}(\theta)$ $(= \widehat{h}_{2,n}(\theta, \cdot, \cdot))$, which is an estimator of $h_{2,F}(\theta, g, g^*)$, is defined by:

$$\begin{aligned} \widehat{h}_{2,n}(\theta, g, g^*) &= \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta, g, g^*)\widehat{D}_n^{-1/2}(\theta), \text{ where} \\ \widehat{\Sigma}_n(\theta, g, g^*) &= n^{-1}\sum_{i=1}^{n} (m(W_i, \theta, g, b) - \overline{m}_n(\theta, g))\,(m(W_i, \theta, g^*, b) - \overline{m}_n(\theta, g^*))' \text{ and} \\ \widehat{D}_n(\theta) &= Diag(\widehat{\Sigma}_n(\theta, 1_k, 1_k)). \end{aligned} \tag{4.6}$$

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

[11] Note that $D_F(\theta, z) = Diag(\sigma_{F,1}^2(\theta, z), ..., \sigma_{F,k}^2(\theta, z))$, where $\sigma_{F,j}^2(\theta, z)$ is defined in (2.14). 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 AN5 and AN6 in Appendix 1.

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

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

## 4.2 GMS Critical Values for Approximate Test Statistics

When the test statistic is approximated via a truncated sum, simulated integral, or QMC quantity, as discussed in Section 3.5, the statistic $T(h)$ in Section 4.1 is replaced by

$$\overline{T}_{s_n}(h) = \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S(\nu_{h_2}(g_\ell) + h_1(g_\ell), h_2(g_\ell, g_\ell) + \varepsilon I_k), \tag{4.7}$$

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

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

$$c_{s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) = c_{0,s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta. \tag{4.8}$$

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

## 4.3 Bootstrap GMS Critical Values

Bootstrap versions of the GMS critical value in (4.4) and the A-GMS critical value in (4.8) can be employed. The bootstrap GMS critical value is

$$c^*(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha) = c^*_0(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{4.9}$$

where $c^*_0(h, 1-\alpha)$ is the $1-\alpha$ quantile of $T^*(h)$ and $T^*(h)$ is defined as in (4.1) but with $\{\nu_{h_2}(g) : g \in \mathcal{G}\}$ and $\widehat{h}_{2,n}(\theta)$ replaced by the bootstrap empirical process $\{\nu^*_n(g) :$

$g \in \mathcal{G}\}$ and the bootstrap covariance kernel $\widehat{h}^*_{2,n}(\theta)$, respectively. By definition, $\nu^*_n(g) = n^{-1/2}\sum_{i=1}^n (m(W^*_i, \theta, g, b) - \overline{m}_n(\theta, g))$, where $\{W^*_i : i \le n\}$ is an i.i.d. bootstrap sample drawn from the empirical distribution of $\{W_i : i \le n\}$. Also, $\widehat{h}^*_{2,n}(\theta, g, g^*)$, $\widehat{\Sigma}^*_n(\theta, g, g^*)$, and $\widehat{D}^*_n(\theta)$ are defined as in (4.6) with $W^*_i$ in place of $W_i$. Note that $\widehat{h}^*_{2,n}(\theta, g, g^*)$ only enters $c(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha)$ via functions $(g, g^*)$ such that $g = g^*$.

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

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

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

$$\xi_n(\theta, g) = \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta, g)\overline{m}_n(\theta, g), \text{ where } \overline{D}_n(\theta, g) = Diag(\overline{\Sigma}_n(\theta, g)), \qquad (4.10)$$

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

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

$$\begin{aligned} \varphi_{n,j}(\theta, g) &= -\eta && \text{if } \xi_{n,j}(\theta, g) \le 1 \\ \varphi_{n,j}(\theta, g) &= \overline{h}_{2,n,j}(\theta, g)^{1/2} B_n && \text{if } \xi_{n,j}(\theta, g) > 1, \text{ where} \\ \overline{h}_{2,n}(\theta, g) &= \widehat{D}_n^{-1/2}(\theta)\overline{\Sigma}_n(\theta, g)\widehat{D}_n^{-1/2}(\theta), \ \overline{h}_{2,n,j}(\theta, g) = [\overline{h}_{2,n}(\theta, g)]_{jj}, && \end{aligned} \qquad (4.11)$$

and $\eta > 0$ is the IUF employed in (4.4).[12]

We assume:

[12] Note that $\varphi_n(\theta, g)$ is defined in AS1 with 0 in place of $-\eta$. The quantity $-\eta$ is required in the definition of $\varphi_n(\theta, g)$ in this paper because it is possible for $E_F m_j(W_i, \theta, g, b)$ to take on small negative values (that converge to 0 as $n \to \infty$) for $j \le p$ because $E_F(m_j(W_i, \theta, g)|Z_i = z) \ge 0$ only holds for $z = z_0$ by (2.1) or (2.15).

**Assumption GMS1.** (a) $\varphi_n(\theta, g)$ satisfies (4.11), where $\{B_n : n \geq 1\}$ is a non-decreasing sequence of positive constants, and

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

The constants $\{B_n : n \geq 1\}$ in Assumption GMS1 need not diverge to infinity for the GMS CS to have asymptotic size greater than or equal to $1 - \alpha$. However, for the GMS CS not to be asymptotically conservative, $B_n$ must diverge to $\infty$, see Assumption GMS2(b) below. In the simulations in Section 8, we use $\kappa_n = (0.3 \ln(n))^{1/2}$ and $B_n = (0.4 \ln(n)/ \ln\ln(n))^{1/2}$, which satisfy Assumption GMS1.

The multiplicand $\overline{h}_{2,n,j}(\theta, g)^{1/2}$ in (4.11) is an "$\varepsilon$-adjusted" standard deviation estimator for the $j$th normalized sample moment based on $g$. It provides a suitable scaling for $\varphi_n(\theta, g)$.

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

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

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

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

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

## 4.5 "Plug-in Asymptotic" Confidence Sets

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 8 show better performance by GMS critical values than PA and subsampling critical values.

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

$$c(-\eta \times 1_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha) = c_0(-\eta \times 1_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{4.12}$$

where $\eta$ is an arbitrarily small positive constant (i.e., an IUF), $1_{\mathcal{G}}$ denotes the $R^k$-valued function on $\mathcal{G}$ that is identically $(1, ..., 1)' \in R^k$, and $\widehat{h}_{2,n}(\theta)$ is defined in (4.6). The nominal $1-\alpha$ PA CS is given by (2.3) with the critical value $c_{n,1-\alpha}(\theta)$ equal to $c(-\eta \times 1_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha)$.

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

# 5 Uniform Asymptotic Coverage Probability Results

In this section, we show that GMS and PA CS's have correct uniform asymptotic coverage probabilities, i.e., correct asymptotic size.

For simplicity, let $h_{2,F}(\theta)$ abbreviate the asymptotic covariance kernel $\{h_{2,F}(\theta, g, g^*) : g, g^* \in \mathcal{G}\}$ defined in (4.5). Define

$$\mathcal{H}_2 = \{h_{2,F}(\theta) : (\theta, F) \in \mathcal{F}\}. \tag{5.13}$$

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

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{g,g^* \in \mathcal{G}} ||h_2^{(1)}(g, g^*) - h_2^{(2)}(g, g^*)||. \tag{5.14}$$

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

**Theorem N1.** *Suppose Assumptions* B, K, NM, S1, *and* S2 *hold and Assumption* GMS1 *also holds when considering* GMS CS's. *Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$, GMS *and* PA *confidence sets* $CS_n$ *satisfy*

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

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

*Assumption* GMS2), *then the* GMS *confidence set satisfies*

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

*where $\eta$ 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 are at most infinitesimally conservative asymptotically. The uniformity results hold whether the moment conditions involve "weak" or "strong" instrumental variables $X_i$.

**2.** As in AS1, an analogue of Theorem N1(b) holds for PA CS's if Assumption GMS2(a) holds and $E_{F_c}(m_j(W_i, \theta_c)|X_i, Z_i = z_0) = 0$ 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.

# 6 Power Against Fixed Alternatives

We now show that the power of GMS and PA tests converges to one as $n \to \infty$ for all fixed alternatives (for which the moment functions have $4+\delta$ moments finite). Thus, both tests are consistent tests. This implies that for any fixed distribution $F_0$ and any parameter value $\theta_*$ *not* in the identified set $\Theta_{F_0}$, the GMS and PA CS's do not include $\theta_*$ 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

$$\begin{aligned} H_0 : & E_{F_0}(m_j(W_i, \theta_*)|X_i, Z_i = z_0) \geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1, ..., p \text{ and} \\ & E_{F_0}(m_j(W_i, \theta_*)|X_i, Z_i = z_0) = 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1, ..., k, \end{aligned} \tag{6.1}$$

where $\theta_*$ 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 $(\theta, F)$ that satisfy conditions (i)-(xiv) in (2.15) that define $\mathcal{F}$ except conditions (iii) and (iv) (which impose the conditional moment inequalities and equalities). As defined, $\mathcal{F} \subset \mathcal{F}_+$. Note that $\mathcal{F}_+$ includes $(\theta, F)$ pairs for which $\theta$ lies outside of the identified set $\Theta_F$ as well as all values in the identified set.

**Assumption NFA.** The 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 (3.9), and (b) $(\theta_*, F_0) \in \mathcal{F}_+$.

Assumption NFA(a) states that violations of the conditional moment inequalities or equalities occur for the null parameter $\theta_*$ for $X_i$ values in some set with positive conditional probability given $Z_i = z_0$ under $F_0$. Thus, under Assumption NFA(a), the moment conditions specified in (6.1) do not hold.

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

$$\begin{aligned} m_j^*(g) &= E_{F_0}(m_j(W_i, \theta_*)g_j(X_i)|Z_i = z_0)f(z_0)/\sigma_{F_0,j}(\theta_*, z_0) \text{ and} \\ \beta(g) &= \max\{-m_1^*(g), ..., -m_p^*(g), |m_{p+1}^*(g)|, ..., |m_k^*(g)|\}. \end{aligned} \tag{6.2}$$

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

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

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

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

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

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

Assumption Q holds for $Q_{AR}$ and $\mathcal{G}_{c\text{-}cube}$ defined above because $\mathcal{G}_{c\text{-}cube}$ is countable and $Q_{AR}$ has a probability mass function that is positive at each element in $\mathcal{G}_{c\text{-}cube}$. Appendix B of AS2 verifies Assumption Q for four other choices of $Q$ and $\mathcal{G}$.

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

**Theorem N2.** *Suppose Assumptions* B, K, NFA, NCI, Q, S1, S3, *and* S4 *hold and* Assumption NM *holds with $F_0$ in place of $F_n$ in* Assumption NM(b). 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.** Assumption NM holds for $\mathcal{G}_{c\text{-}cube}$ *with* $F_0$ *in place of* $F_n$ in part (b) because $\mathcal{C}_{c\text{-}cube}$ is a Vapnik-Cervonenkis class of sets.

# 7 Power Against $(\mathbf{nb}^{\mathbf{d_z}})^{-1/2}$-Local Alternatives

In this section, we show that GMS and PA tests have power against certain, but not all, $(nb^{d_z})^{-1/2}$-local alternatives. These testing results have immediate implications for the volume of CS's, see Pratt (1961).

We show that a GMS test has asymptotic power that is greater than or equal to that of a PA test (based on the same test statistic) under all alternatives with strict inequality in certain scenarios.

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

$$\begin{aligned} H_0 : & E_{F_n}(m_j(W_i, \theta_{n,*})|Z_i = z_0) \geq 0 \text{ for } j = 1, ..., p, \\ & E_{F_n}(m_j(W_i, \theta_{n,*})|Z_i = z_0) = 0 \text{ for } j = p+1, ..., k, \end{aligned} \tag{7.1}$$

and $(\theta_{n,*}, F_n) \in \mathcal{F}$, where $F_n$ denotes the true distribution of the data. The null values $\theta_{n,*}$ are allowed to drift with $n$ or be fixed for all $n$. Drifting $\theta_{n,*}$ values are of interest because they allow one to consider the case of a fixed identified set, say $\Theta_0$, and to derive the asymptotic probability that parameter values $\theta_{n,*}$ that are not in the identified set, but drift toward it at rate $n^{-1/2}$, are excluded from a GMS or PA CS. In this scenario, the sequence of true distributions are ones that yield $\Theta_0$ to be the identified set, i.e., $F_n \in \mathcal{F}_0 = \{F : \Theta_F = \Theta_0\}$.

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

Let $f_n(z)$ denote the density of $Z_i$ wrt $\mu_{Leb}$ under $F_n$.

The $(nb^{d_z})^{-1/2}$-local alternatives are defined as follows.

**Assumption NLA1.** The true parameters and distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ and the null parameters $\{\theta_{n,*} : n \geq 1\}$ satisfy:

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

(b) $(nb^{d_z})^{1/2}E_{F_n}(m_j(W_i,\theta_n,g)|Z_i=z_0)f_n(z_0)/\sigma_{F_n,j}(\theta_n,z_0)\to h_{1,j}(g)$ for some $h_{1,j}(g)$ $\in R_{+,\infty}$ for $j=1,...,p$ and all $g\in\mathcal{G}$,

(c) $d(h_{2,F_n}(\theta_n),h_{2,F_0}(\theta_0))\to 0$ and $d(h_{2,F_n}(\theta_{n,*}),h_{2,F_0}(\theta_0))\to 0$ as $n\to\infty$ (where $d$ is defined in (5.14)), and

(d) $(\theta_n,F_n)\in\mathcal{F}_+$ for all $n\geq 1$.

**Assumption NLA2.** The $k\times d$ matrix $\Pi_F(\theta,g)=(\partial/\partial\theta')[D_F^{-1/2}(\theta,z_0)E_F(m(W_i,\theta,g)$ $|Z_i=z_0)f(z_0)]$ exists and is continuous in $(\theta,F)$ for all $(\theta,F)$ in a neighborhood of $(\theta_0,F_0)$ for all $g\in\mathcal{G}$.

For notational simplicity, we let $h_2$ abbreviate $h_{2,F_0}(\theta_0)$ throughout this section. Assumption NLA1(a) states that the true values $\{\theta_n:n\geq 1\}$ are $(nb^{d_z})^{-1/2}$-local to the null values $\{\theta_{n,*}:n\geq 1\}$. Assumption NLA1(b) specifies the asymptotic behavior of the (normalized) moment inequality functions when evaluated at the true values $\{\theta_n:n\geq 1\}$. Under the true values, these (normalized) moment inequality functions are non-negative. Assumption NLA1(c) specifies the asymptotic behavior of the covariance kernels $\{h_{2,F_n}(\theta_n,\cdot,\cdot):n\geq 1\}$ and $\{h_{2,F_n}(\theta_{n,*},\cdot,\cdot):n\geq 1\}$. Assumption NLA2 is a smoothness condition on the normalized expected conditional moment functions given $Z_i=z_0$. Given the smoothing properties of the expectation operator, this condition is not restrictive.

Under Assumptions NLA1 and NLA2, we show that the moment inequality functions evaluated at the null values $\{\theta_{n,*}:n\geq 1\}$ satisfy:

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

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

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

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

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

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

Assumption LA3$'$ requires that either (i) the moment equalities detect violations of the null hypothesis for a set of $g$ functions with positive $Q$ measure or (ii) the moment inequalities are not too far from being binding, i.e., $h_{1,j}(g) < \infty$, and the perturbation $\lambda_0$ occurs in a direction that yields moment inequality violations for a set of $g$ functions with positive $Q$ measure.

Assumption LA3 is employed with the KS test. It is weaker than Assumption LA3$'$. If Assumption LA3 holds with $\lambda = \beta\lambda_0$ (and some other assumptions), then the power of KS-GMS and KS-PA tests against the perturbation $\lambda$ is arbitrarily close to one for $\beta$ arbitrarily large. For brevity, we do not prove this here. The proof is analogous to the proof of such results for the KS tests considered in AS1 and AS2, see Section 13.1 of Appendix B and Section 15.1 of Appendix E of AS2.

Assumptions LA3 and LA3$'$ can fail to hold even when the null hypothesis is violated. This typically happens if the true parameter/true distribution is fixed, i.e., $(\theta_n, F_n) = (\theta_0, F_0) \in \mathcal{F}$ for all $n$ in Assumption NLA1(a), the null hypothesis parameter $\theta_{n,*}$ drifts with $n$ as in Assumption NLA1(a), and $P_{F_0}(X_i \in \mathcal{X}_{zero}|Z_i = z_0) = 0$, where $\mathcal{X}_{zero} = \{x \in R^{d_x} : E_{F_0}(m(W_i, \theta_0)|X_i = x, Z_i = z_0) = 0\}$. In such cases, typically $h_{1,j}(g) = \infty$ $\forall g \in \mathcal{G}$ (because the conditional moment inequalities are non-binding with probability one), Assumptions LA3 and LA3 $'$ fail, and Theorem N3 below shows that GMS and PA tests have trivial asymptotic power against these $(nb^{d_z})^{-1/2}$-local alternatives.

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

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

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

Next, we introduce two assumptions, viz., Assumptions NLA4 and LA5, that are used only for GMS tests in the context of local alternatives. The asymptotic analogues of $\overline{\Sigma}_n(\theta, g)$ and its diagonal matrix are

$$\overline{\Sigma}_F(\theta, g, z_0) = \Sigma_F(\theta, g, g, z_0) + \varepsilon\Sigma_F(\theta, 1_k, 1_k, z_0) \text{ and } \overline{D}_F(\theta, g, z_0) = Diag(\overline{\Sigma}_F(\theta, g, z_0)), \tag{7.4}$$

where $\Sigma_F(\theta, g, g, z_0)$ is defined in (4.5).

**Assumption NLA4.** $\kappa_n^{-1}(nb^{d_z})^{1/2}\overline{D}_{F_n}^{-1/2}(\theta_n, g, z_0)E_{F_n}(m(W_i, \theta_n, g)|Z_i = z_0)f(z_0) \to \pi_1(g)$, where $\pi_1(g) = (\pi_{1,1}(g), ..., \pi_{1,k}(g))'$, for some $\pi_{1,j}(g) \in R_{+,\infty}$ for $j = 1, ..., p$, $\pi_{1,j}(g) = 0$ for $j = p+1, ..., k$, and all $g \in \mathcal{G}$.

In Assumption NLA4 the functions are evaluated at the true value $\theta_n$, not at the null value $\theta_{n,*}$, and $(\theta_n, F_n) \in \mathcal{F}$. In consequence, the moment functions in Assumption NLA4 satisfy the moment inequalities and $\pi_{1,j}(g) \geq 0$ for all $j = 1, ..., p$ and $g \in \mathcal{G}$. Note that $0 \leq \pi_{1,j}(g) \leq h_{1,j}(g)$ for all $j = 1, ..., p$ and all $g \in \mathcal{G}$ (by Assumption NLA1(b) and $\kappa_n \to \infty$.)

Let $c_0(\varphi(\pi_1), h_2, 1-\alpha)$ denote the $1-\alpha$ quantile of

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

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

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

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

The value 1 that appears in $\mathcal{G}_\varphi$ in Assumption LA5(a) is the discontinuity point of $\varphi$. Assumption LA5(a) implies that the $(nb^{d_z})^{-1/2}$-local power formulae given below do not apply to certain "discontinuity vectors" $\pi_1(\cdot)$, but this is not particularly restrictive.[13]

[13] Assumption LA5(a) is not particularly restrictive because in cases where it fails, one can obtain lower and upper bounds on the local asymptotic power of GMS tests by replacing $c(\varphi(\pi_1), h_2, 1-\alpha)$ by $c(\varphi(\pi_1-), h_2, 1-\alpha)$ and $c(\varphi(\pi_1+), h_2, 1-\alpha)$, respectively, in Theorem N3(a). By definition, $\varphi(\pi_1-) = \varphi(\pi_1(\cdot)-)$ and $\varphi(\pi_1(g)-)$ is the limit from the left of $\varphi(x)$ at $x = \pi_1(g)$. Likewise $\varphi(\pi_1+) = \varphi(\pi_1(\cdot)+)$ and $\varphi(\pi_1(g)+)$ is the limit from the right of $\varphi(x)$ at $x = \pi_1(g)$.

Assumption LA5(b) typically holds because of the absolute continuity of the Gaussian random variables $\nu_{h_2}(g)$ that enter $T(\varphi(\pi_1), h_2)$.[14]

The following assumption is used only for PA tests.

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

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

**Theorem N3.** *Under Assumptions* B, K, NM, S1, S2, *and* NLA1-NLA2,

(a) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(\varphi_n(\theta_{n,*}), \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)) = 1 - J_{h,\lambda}(c(\varphi(\pi_1), h_2, 1-\alpha))$ *provided Assumptions* GMS1, NLA4, *and* LA5 *also hold,*

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

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

**Comments. 1.** Theorems N3(a) and N3(b) provide the $(nb^{d_z})^{-1/2}$-local alternative power functions of the GMS and PA tests, respectively. Theorem N3(c) shows that the asymptotic power of GMS and PA tests is arbitrarily close to one if the $(nb^{d_z})^{-1/2}$-local alternative parameter $\lambda = \beta\lambda_0$ is sufficiently large in the sense that its scale $\beta$ is large.

**2.** We have $c(\varphi(\pi_1), h_2, 1-\alpha) \leq c(0_{\mathcal{G}}, h_2, 1-\alpha)$ (because $\varphi(\pi_1(g)) \geq 0$ for all $g \in \mathcal{G}$ and $S(m, \Sigma)$ is non-increasing in $m_I$ by Assumption S1(b), where $m = (m_I', m_{II}')'$). Hence, the asymptotic local power of a GMS test is greater than or equal to that of a PA test. Strict inequality holds whenever $\pi_1(\cdot)$ is such that $Q(\{g \in \mathcal{G} : \varphi(\pi_1(g)) > 0\}) > 0$. The latter typically occurs whenever the conditional moment inequality $E_{F_n}(m_j(W_i, \theta_{n,*}) |X_i, Z_i = z_0)$ for some $j = 1, ..., p$ is bounded away from zero as $n \to \infty$ with positive $X_i$ probability.

**3.** The results of Theorem N3 hold under the null hypothesis as well as under the alternative. The results under the null quantify the degree of asymptotic non-similarity of the GMS and PA tests.

[14] If Assumption LA5(b) fails, one can obtain lower and upper bounds on the local asymptotic power of GMS tests by replacing $J_{h,\lambda}(c(\varphi(\pi_1),\ h_2, 1-\alpha))$ by $J_{h,\lambda}(c(\varphi(\pi_1),\ h_2, 1-\alpha)+)$ and $J_{h,\lambda}(c(\varphi(\pi_1),\ h_2, 1-\alpha)-)$, respectively, in Theorem N3(a), where the latter are the limits from the left and right, respectively, of $J_{h,\lambda}(x)$ at $x = c(\varphi(\pi_1), h_2, 1-\alpha)$.

**4.** Suppose the assumptions of Theorem N3 hold and each distribution $F_n$ generates the same identified set, call it $\Theta_0 = \Theta_{F_n}$ $\forall n \geq 1$. Then, Theorem N3(a) implies that the asymptotic probability that a GMS CS includes, $\theta_{n,*}$, which lies within $O((nb^{d_z})^{-1/2})$ of the identified set, is $J_{h,\lambda}(c(\varphi(\pi_1), h_2, 1-\alpha))$. If $\lambda = \beta\lambda_0$ and Assumptions LA3$'$, S3, and S4 also hold, then $\theta_{n,*}$ is not in $\Theta_0$ (at least for $\beta$ large) and the asymptotic probability that a GMS or PA CS includes $\theta_{n,*}$ is arbitrarily close to zero for $\beta$ arbitrarily large by Theorem N3(c). Analogous results hold for PA CS's.

# 8 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.

## 8.1 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 Sections 3 and 4. 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.[15] The critical values are simulated using 5001 repetitions (for each original sample repetition). The base case values of $\kappa_n$, $B_n$, and $\varepsilon$ 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.

A bandwidth $b$ and a kernel function are required to compute the test statistic and the critical values. The kernel function is chosen to be the Epanechnikov kernel: $K(x) = 0.75\max\{1 - x^2, 0\}$. We use the bandwidth $b = b^0 n^{-2/7}$, where $b^0 = 4.68\hat{\sigma}_z$ and $\hat{\sigma}_z$ is the estimated standard deviation of $Z_i$.[16] Both the kernel function and the bandwidth selection procedure are the same for both simulation examples. For comparative purposes, some results are also reported for $b = 0.5b^0 n^{-2/7}$ and $b = 2b^0 n^{-2/7}$.

## 8.2 Nonparametric Quantile Selection

This model extends the quantile selection model in AS1. We are interested in the conditional $\tau$-quantile of a treatment response given the value of covariates $X_i$ and $Z_i$.

[15] 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 $\eta$ 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.

[16] 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.

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, $\theta$, is the conditional $\tau$-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 $\tau$-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:

$$\begin{aligned} E\left(1(X_i \leq x_0)[1(Y_i \leq \theta, T_i = t) + 1(T_i \neq t) - \tau]|X_i, Z_i = z_0\right) &\geq 0 \text{ a.s. and} \\ E\left(1(X_i \geq x_0)[\tau - 1(Y_i \leq \theta, T_i = t)]|X_i, Z_i = z_0\right) &\geq 0 \text{ a.s.} \end{aligned} \tag{8.1}$$

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

$$\begin{aligned} y_i(1) &= \mu(X_i, Z_i) + \sigma\left(X_i, Z_i\right) u_i, \text{ where } \partial\mu\left(x, z\right)/\partial x \geq 0 \text{ and } \sigma\left(x, z\right) \geq 0, \\ T_i &= 1\{L\left(X_i, Z_i\right) + \varepsilon_i \geq 0\}, \text{ where } \partial L\left(x, z\right)/\partial x \geq 0, \\ X_i, Z_i &\sim Unif[0, 2], \; (\varepsilon_i, u_i) \sim N(0, I_2), \; (X_i, Z_i) \perp (\varepsilon_i, u_i), \; X_i \perp Z_i, \\ Y_i &= y_i(T_i), \text{ and } t = 1. \end{aligned} \tag{8.2}$$

The variable $y_i(0)$ is irrelevant (because $Y_i$ enters the moment inequalities in (8.1) 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 (8.1) imply:

$$\theta \geq \underline{\theta}(x, z_0) := \mu(x, z_0) + \sigma(x, z_0)\,\Phi^{-1}\left(1 - [2\Phi(L(x, z_0))]^{-1}\right) \text{ for } x \leq 1.5 \text{ and}$$
$$\theta \leq \bar{\theta}(x, z_0) := \mu(x, z_0) + \sigma(x, z_0)\,\Phi^{-1}\left([2\Phi(L(x, z_0))]^{-1}\right) \text{ for } x \geq 1.5. \tag{8.3}$$

We call $\underline{\theta}(x, z_0)$ and $\bar{\theta}(x, z_0)$ the lower and upper bound functions on $\theta$, respectively. The identified set for the quantile selection model is $\left[\sup_{x \leq x_0} \underline{\theta}(x, z_0),\ \inf_{x \geq x_0} \bar{\theta}(x, z_0)\right]$. The shape of the lower and upper bound functions depends on the $\mu$, $\sigma$, 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.[17]

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

$$\beta(x, z, \theta) = \begin{cases} E\left(1(Y_i \leq \theta, T_i = 1) + 1(T_i \neq 1) - 0.5 \mid (X_i, Z_i) = (x, z)\right) & \text{if } x < 1.5 \\ E\left(0.5 - 1(Y_i \leq \theta, T_i = 1) \mid (X_i, Z_i) = (x, z)\right) & \text{if } x \geq 1.5. \end{cases} \tag{8.4}$$

The conditional moment functions as functions of $x$ at $z = z_0$ are flat, kinked and peaked under the three specifications of $\mu$, $\sigma$, 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.

### 8.2.1 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 $(0, 1)$.[18] The hypercubes have side-edge lengths

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

[18] 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.

$(2r)^{-1}$ for $r = r_0, ..., r_1$, where $r_0 = 1$ and the base case value of $r_1$ is 3.[19] 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 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.

#### 8.2.2 Simulation Results

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.[20]

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 $(\kappa_n, B_n)$ for the GMS/Asy critical values, (iv) the value of $\varepsilon$, 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

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

[20] Note that the DGP is the same for FCP's as for CP's, just the value $\theta$ that is to be covered is different. For the lower endpoint of the identified set, FCP's are computed for $\theta$ equal to $\sup_{x \le 1.5} \underline{\theta}(x, 1) - c \times (250/n)^{5/14}$, 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.

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,

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

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | (a) Coverage Probabilities | | | | | |
| | Statistic: | CvM/ Sum | CvM/ QLR | CvM/ Max | KS/ Sum | KS/ QLR | KS/ Max |
| DGP | Crit Val | | | | | | |
| 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) | | | | | |
| 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 |

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 versions of the CI's regardless of the DGP specification or the test statistic choice (CvM

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

| | | | | | | |
|---|---|---|---|---|---|---|
| | | (a) Coverage Probabilities | | | | |
| | Critical Value: | PA/Asy | PA/Bt | GMS/Asy | GMS/Bt | Sub |
| DGP | Statistic | | | | | |
| 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) | | | | |
| 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 |

or KS).[21]

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 |
| $b = 0.5 b^0 n^{-2/7}$ | | .960 | .963 | .68 | .77 |
| $b = 2 b^0 n^{-2/7}$ | | .950 | .948 | .19 | .34 |
| $\alpha = .5$ | | .525 | .516 | .045 | .072 |
| $\alpha = .5$ & $n = 500$ | | .517 | .519 | .042 | .070 |

[21] 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.

with a smaller $\varepsilon$.[22] There is some sensitivity to the magnitude of the GMS tuning parameters $(\kappa_n, B_n)$—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 bound as shown in Tables A1 and A2 in Appendix 2 in Andrews and Shi (2010a).

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 $n = 250$.

To sum up, we find that the CI's based on the CvM /Max statistic with the GMS/Asy critical value perform the best 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.

## 8.3 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

[22] The $\theta$ 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.

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:

$$E[y_i(1) - y_i(0)|(X_i, Z_i) = (x, z_0)] \geq 0 \text{ for all } x \in \mathcal{X}. \tag{8.5}$$

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 $\Pr(D_i = 1) = p \in (0, 1)$.[23] Then,

$$E[y_i(1) - y_i(0)|(X_i, Z_i) = (x, z_0)] = E\left[\frac{Y_i D_i}{p} - \frac{Y_i(1 - D_i)}{1 - p}|(X_i, Z_i) = (x, z_0)\right]. \tag{8.6}$$

Then, the hypothesis (8.5) is equivalent to testing if $\theta = 0$ is in the identified set of the following moment inequality model:

$$E\left[\frac{Y_i D_i}{p} - \frac{Y_i(1 - D_i)}{1 - p} - \theta|(X_i, Z_i) = (x, z_0)\right] \geq 0 \text{ for all } x \in \mathcal{X}. \tag{8.7}$$

[23] 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).

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

$$
\begin{aligned}
y_i(0) &= 0,\ y_i(1) = \mu(X_i, Z_i) + u_i,\ D_i = 1\{\varepsilon_i \geq 0\}, \\
X_i &\sim Unif[0,2],\ Z_i \sim Unif[-1,1],\ (\varepsilon_i, u_i) \sim N(0, I_2), \\
(X_i, Z_i) &\perp (\varepsilon_i, u_i),\ \text{and } X_i \perp Z_i.
\end{aligned} \tag{8.8}
$$

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 (8.5) holds if $a = 0$ and is violated if $a > 0$. The functions $\mu_1$ and $\mu_2$ do not change sign in a neighborhood around $z_0$, whereas the tilted function $\mu_3$ 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.

### 8.3.1 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.

### 8.3.2 Simulation Results

Tables IV and V 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 $a > 0$.[24]

Table IV provides comparisons of the PA/Asy, PA/Bt, GMS/Asy, GMS/Bt, and Sub

[24] 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.

critical values for the CvM and KS test statistics. Table V provides robustness results for the CvM and KS test statistics in the flat bound case. Table V 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 $(\kappa_n, B_n)$ for the GMS/Asy critical values, (iv) the value of $\varepsilon$, upon which the variance estimator $\overline{\Sigma}_n(\theta, g)$ depends, and (v) the bandwidth $b$.

Table IV 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

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

| | | | | | | |
|---|---|---|---|---|---|---|
| | (a) Null Rejection Probabilities | | | | | |
| | 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 |

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 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 IV shows (i) a clear advantage of CvM-based tests 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 V 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 Appendix 2 (in Andrews and Shi (2010a)). The results in Table V show little sensitivity to the sample size and a smaller $\varepsilon$ 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 $(\kappa_n, B_n)$. Increasing the number of cubes or increasing $(\kappa_n, B_n)$ 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 Appendix 2 (in Andrews and Shi (2010a)).

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.

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

| | | (a) Null Rejection Probabilities | | (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 |
| $b = 0.5 b^0 n^{-2/7}$ | | .041 | .020 | .28 | .14 |
| $b = 2 b^0 n^{-2/7}$ | | .049 | .043 | .78 | .57 |

# 9 Appendix 1

## 9.1 S Function Assumptions

Let $m_I = (m_1, ..., m_p)'$ and $m_{II} = (m_{p+1}, ..., m_k)'$. Let $\Delta$ be the set of $k \times k$ positive-definite diagonal matrices. Let $\mathcal{W}$ be the set of $k \times k$ positive-definite matrices. Let $\mathcal{S} = \{(m, \Sigma) : m \in R^p_{[+\infty]} \times R^v,\ \Sigma \in \mathcal{W}\}$. Let $R_+ = \{x \in R : x \geq 0\}$.

We consider functions $S$ that satisfy the following conditions.

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

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

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

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

(d) $S$ is continuous, and

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

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

**Assumption S2.** $S(m, \Sigma)$ is uniformly continuous in the sense that, for all $m_0 \in R^k$ and all pd $\Sigma_0$, $\sup_{\mu \in R^p_+ \times \{0\}^v} |S(m + \mu, \Sigma) - S(m_0 + \mu, \Sigma_0)| \to 0$ as $(m, \Sigma) \to (m_0, \Sigma_0)$.[25]

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

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

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

## 9.2 X-Instrument Assumptions

The collection $\mathcal{G}$ must satisfy the following "manageability" condition. The manageability condition is from Pollard (1990) and is defined and explained in Appendix E of AS2.

[25] Assumption S2 is equivalent to the same condition with $\mu$ vectors whose elements exceed $-\eta_1$ for some $\eta_1 < \infty$. This is used in the proofs below.

**Assumption NM.** (a) $0 \leq g_j(x) \leq G \ \forall x \in R^{d_x}, \forall j \leq k, \forall g \in \mathcal{G}$, for some constant $G < \infty$, and

(b) the processes $\{g_j(X_{n,i}) : g \in \mathcal{G},\ i \leq n, n \geq 1\}$ are manageable with respect to the constant function $G$ for $j = 1, ..., k$, where $\{X_{n,i} : i \leq n, n \geq 1\}$ is a row-wise i.i.d. triangular array with $X_{n,i} \sim F_{X,n}$ and $F_{X,n}$ is the distribution of $X_{n,i}$ under $F_n$ for some $(\theta_n, F_n) \in \mathcal{F}_+$ for $n \geq 1$.[26]

## 9.3 Proof of Lemma N1

**Proof of Lemma N1.** We have: $\theta \notin \Theta_F(\mathcal{G})$ implies that $E_F(m_j(W_i, \theta)g_j(X_i)|Z_i = z_0) < 0$ for some $j \leq p$ or $E_F(m_j(W_i, \theta)g_j(X_i)|Z_i = z_0) \neq 0$ for some $j = p+1, ..., k$. By the law of iterated expectations and $g_j(x) \geq 0$ for all $x \in R^{d_x}$ and $j \leq p$, this implies that $P_F(X_i \in \mathcal{X}_F(\theta)|Z_i = z_0) > 0$ and, hence, $\theta \notin \Theta_F$.

On the other hand, $\theta \notin \Theta_F$ implies that $P_F(X_i \in \mathcal{X}_F(\theta)|Z_i = z_0) > 0$ and the latter implies that $\theta \notin \Theta_F(\mathcal{G})$ by Assumption NCI. $\square$

## 9.4 Proof of Theorem N1

In this section, we prove Theorem N1. We start by introducing some notation. Next, we establish Theorem AN1, which is used in the proof of Theorem N1. To prove Theorem AN1 we use Lemmas AN1-AN3. The proofs of the latter use Lemmas AN4-AN6.

### 9.4.1 Notation

First, we define sample-size $n$ population analogues of the asymptotic covariance kernels that are defined in (4.5). We make their dependence on $b = b_n$ explicit. Let[27]

$$\begin{aligned} h_{2,F}(\theta, g, g^*, b) &= D_F^{-1/2}(\theta, b)\Sigma_F(\theta, g, g^*, b)D_F^{-1/2}(\theta, b) \\ &= Cov_F\left(D_F^{-1/2}(\theta, b)m(W_i, \theta, g, b), D_F^{-1/2}(\theta, b)m(W_i, \theta, g^*, b)\right), \\ \Sigma_F(\theta, g, g^*, b) &= Cov_F(m(W_i, \theta, g, b), m(W_i, \theta, g^*, b)), \text{ and} \qquad (9.1) \\ D_F(\theta, b) &= Diag(\Sigma_F(\theta, 1_k, 1_k, b)) \ (= Diag(Var_F(b^{-d_z/2}K_b(Z_i)m(W_i, \theta)))). \end{aligned}$$

[26] The asymptotic results given in the paper hold with Assumption NM replaced by any alternative assumption that is sufficient to obtain the requisite empirical process results given in Lemma AN3 below.

[27] For simplicity, there is some abuse of notation in the definitions in (9.1) because $h_{2,F}(\theta, g, g^*, b)$ has a different definition than $h_{2,F}(\theta, g, g^*, z_0)$ in (4.5), but the only difference in the notation is $b$ versus $z_0$. The same is true for $\Sigma_F(\theta, g, g^*, b)$ and $D_F(\theta, b)$ versus $\Sigma_F(\theta, g, g^*, z_0)$ and $D_F(\theta, z_0)$.

Let $h_{2,F}(\theta,b)$ abbreviate the sample-size $n$ covariance kernel $\{h_{2,F}(\theta,g,g^*,b) : g,g^* \in \mathcal{G}\}$ of $n^{1/2}\overline{m}_n(\theta,g)$, which depends on $n$ through $b$.

Next, define

$$\begin{aligned} h_{1,n,F}(\theta,g,b) &= n^{1/2}D_F^{-1/2}(\theta,b)E_F m(W_i,\theta,g,b), \\ h_{n,F}(\theta,g,g^*,b) &= (h_{1,n,F}(\theta,g,b), h_{2,F}(\theta,g,g^*,b)), \\ \widehat{h}_{2,n,F}(\theta,g,g^*,b) &= D_F^{-1/2}(\theta,b)\widehat{\Sigma}_n(\theta,g,g^*)D_F^{-1/2}(\theta,b), \\ \overline{h}_{2,n,F}(\theta,g,b) &= \widehat{h}_{2,n,F}(\theta,g,g,b) + \varepsilon\widehat{h}_{2,n,F}(\theta,1_k,1_k,b) \\ &= D_F^{-1/2}(\theta,b)\overline{\Sigma}_n(\theta,g)D_F^{-1/2}(\theta,b), \text{ and} \\ \nu_{n,F}(\theta,g,b) &= n^{-1/2}\sum_{i=1}^n D_F^{-1/2}(\theta,b)[m(W_i,\theta,g,b) - E_F m(W_i,\theta,g,b)], \end{aligned} \tag{9.2}$$

where $m(W_i,\theta,g,b)$, $\overline{\Sigma}_n(\theta,g)$, and $\widehat{\Sigma}_n(\theta,g,g^*)$ are defined in (3.3), (3.5), and (4.6), respectively. Below we write $T_n(\theta)$ as a function of the quantities in (9.2). As defined, (i) $h_{1,n,F}(\theta,g,b)$ is a $k$-vector of normalized means of the moment functions $D_F^{-1/2}(\theta,b)m(W_i,\theta,g,b)$ for $g \in \mathcal{G}$, which measure the slackness of the population moment conditions under $(\theta,F)$, (ii) $h_{n,F}(\theta,g,g^*,b)$ contains the normalized means of $D_F^{-1/2}(\theta,b)m(W_i,\theta,g,b)$ and the covariances of $D_F^{-1/2}(\theta,b)m(W_i,\theta,g,b)$ and $D_F^{-1/2}(\theta,b)$ $m(W_i,\theta,g^*,b)$, (iii) $\widehat{h}_{2,n,F}(\theta,g,g^*,b)$ and $\overline{h}_{2,n,F}(\theta,g,b)$ are hybrid quantities—part population, part sample—based on $\widehat{\Sigma}_n(\theta,g,g^*)$ and $\overline{\Sigma}_n(\theta,g)$, respectively, and (iv) $\nu_{n,F}(\theta,g,b)$ is the sample average of $D_F^{-1/2}(\theta,b) \times m(W_i,\theta,g,b)$ normalized to have mean zero and variance that is $O(1)$ but not $o(1)$. Note that $\nu_{n,F}(\theta,\cdot,b)$ is an empirical process indexed by $g \in \mathcal{G}$ with covariance kernel given by $h_{2,F}(\theta,g,g^*,b)$.

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

$$n^{1/2}\overline{m}_n(\theta,g) = D_F^{1/2}(\theta,b)(\nu_{n,F}(\theta,g,b) + h_{1,n,F}(\theta,g,b)). \tag{9.3}$$

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

$$T_n(\theta) = \int S(\nu_{n,F}(\theta,g,b) + h_{1,n,F}(\theta,g,b), \overline{h}_{2,n,F}(\theta,g,b))dQ(g). \tag{9.4}$$

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

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

For notational simplicity, for any function of the form $r_F(\theta,g,b)$ for $g \in \mathcal{G}$, let

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

### 9.4.2 Theorem AN1

The following Theorem provides a uniform asymptotic distributional result for the test statistic $T_n(\theta)$. It is an analogue of Theorem 1 of AS1. It used in the proof of Theorem N1.

**Theorem AN1.** *Suppose Assumptions* B, K, NM, S1, *and* S2 *hold. Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$*, all constants* $x_{h_{n,F}(\theta,b)} \in R$ *that may depend on* $(\theta, F)$ *and* $n$ *through* $h_{n,F}(\theta, b)$*, and all* $\delta > 0$*, we have*

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

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

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

**Comments. 1.** Theorem AN1 gives a uniform asymptotic approximation to the distribution function of $T_n(\theta)$. Uniformity holds without any restrictions on the normalized mean (i.e., moment inequality slackness) functions $\{h_{1,n,F_n}(\theta_n, b) : n \geq 1\}$. In particular, Theorem AN1 does not require $\{h_{1,n,F_n}(\theta_n, b) : n \geq 1\}$ to converge as $n \to \infty$ or to belong to a compact set. The Theorem does not require that $T_n(\theta)$ has a unique asymptotic distribution under any sequence $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$.

**2.** The supremum and infimum in Theorem AN1 are over compact sets of asymptotic covariance kernels $\mathcal{H}_{2,cpt}$, rather than the parameter spaces $\mathcal{H}_2$ of covariance kernels. This is not particularly problematic because the potential asymptotic size problems that arise in moment inequality models are due to the pointwise discontinuity of the asymptotic distribution of the test statistic as a function of the means of the moment inequality functions, not as a function of the covariances between different moment inequalities.

### 9.4.3 Lemmas AN1-AN3

The proof of Theorem AN1 uses the following three Lemmas. The first Lemma is a key result that establishes that the finite-sample covariance kernel $h_{2,F}(\theta, b)$ converges to the asymptotic covariance kernel $h_{2,F}(\theta)$ in the sup norm $d$ uniformly over $(\theta, F) \in \mathcal{F}_+$.

**Lemma AN1.** *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$
\begin{aligned}
&\text{(a)} \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g,g^*\in\mathcal{G}} \|\Sigma_F(\theta, g, g^*, b) - \Sigma_F(\theta, g, g^*, z_0)\| \to 0,\\
&\text{(b)} \sup_{(\theta,F)\in\mathcal{F}_+} \left\|D_F^{-1}(\theta, z_0) D_F(\theta, b) - I_k\right\| \to 0, \text{ and}\\
&\text{(c)} \sup_{(\theta,F)\in\mathcal{F}_+} d(h_{2,F}(\theta, b), h_{2,F}(\theta)) \to 0.
\end{aligned}
$$

**Comment.** Lemma AN1 is a key ingredient in the proof of Lemma AN3, which in turn is used in the proofs of Theorems AN1 and N1. See Comment 3 to Lemma AN3 for a description of how Lemma AN1 is employed.

The next Lemma shows that the bias due to taking averages over values $z$ ($\neq z_0$) for which the conditional moment inequalities in (2.1) do *not* hold is negligible asymptotically.

**Lemma AN2.** *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$
\liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}} \inf_{g\in\mathcal{G}} h_{1,n,F}(\theta, g, b) + \eta_1 \geq 0 \ \forall \eta_1 > 0.
$$

**Comment.** Lemma AN2 only applies for $(\theta, F) \in \mathcal{F}$, not $(\theta, F) \in \mathcal{F}_+$.

The next Lemma is analogous to Lemma A1 of AS2. It is used in the proofs of Theorems AN1 and N1-N3. It establishes a functional CLT and uniform LLN for certain independent non-identically distributed empirical processes as well as uniform convergence of the estimator of the covariance kernel.

Let $\mathcal{H}_{2,+} = \{h_{2,F}(\theta) : (\theta, F) \in \mathcal{F}_+\}$. By definition, $\mathcal{H}_{2,+}$ is a set of $k \times k$-matrix-valued covariance kernels on $\mathcal{G} \times \mathcal{G}$ that includes $\mathcal{H}_2$.

**Definition SubSeq($\mathbf{h_2}$).** For $h_2 \in \mathcal{H}_{2,+}$, $SubSeq(h_2)$ is the set of subsequences $\{(\theta_{a_n},$

$F_{a_n}) \in \mathcal{F}_+ : n \geq 1\}$, where $\{a_n : n \geq 1\}$ is some subsequence of $\{n\}$, for which

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

and (ii) $\{W_i : i \geq 1\}$ are i.i.d. under $F_{a_n}$.

Note that the definition of $SubSeq(h_2)$ here differs from the definition of $SubSeq(h_2)$ in AS2 because (i) the summands of the sample averages are $m(W_i, \theta, g, b) = b^{-d_z/2} K_b(Z_i)$ $m(W_i, \theta, g)$, rather than $m(W_i, \theta, g)$, and $\{m(W_i, \theta, g, b) m(W_i, \theta, g^*, b)' : n \geq 1\}$ is not uniformly integrable, which complicates the proof of Lemma AN3(b) below, (ii) $SubSeq$ $(h_2)$ requires $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}_+$, and (iii) $SubSeq(h_2)$ does not impose any conditions related to Assumption NM. The latter are imposed separately in the results below.

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

$$\rho^2(g, g^*) = tr\left(h_2(g, g) - h_2(g, g^*) - h_2(g^*, g) + h_2(g^*, g^*)\right). \tag{9.5}$$

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

$$\begin{aligned} \rho^2(g, g^*) &= E_F ||D_F^{-1/2}(\theta)[\widetilde{m}(W_i, \theta, g) - \widetilde{m}(W_i, \theta, g^*)]||^2, \text{ where} \\ \widetilde{m}(W_i, \theta, g) &= m(W_i, \theta, g) - E_F m(W_i, \theta, g). \end{aligned} \tag{9.6}$$

**Lemma AN3.** *Suppose Assumptions* B *and* NM *hold. For any subsequence* $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$ with $h_2 \in \mathcal{H}_{2,+}$,
(a) $\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot, b_{a_n}) \Rightarrow \nu_{h_2}(\cdot)$ *as* $n \to \infty$ (*as processes indexed by* $g \in \mathcal{G}$), *and*
(b) $\sup_{g,g^* \in \mathcal{G}} ||\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g^*, b_{a_n}) - h_2(g, g^*)|| \to_p 0$ *as* $n \to \infty$.

**Comments. 1.** To obtain uniform asymptotic coverage probability results for CS's, Lemma AN3 is applied with $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ and $h_2 \in \mathcal{H}_2$. To obtain power results under fixed and local alternatives, Lemma AN3 is applied with $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}_+/\mathcal{F}$ for all $n \geq 1$ and $h_2 \in \mathcal{H}_{2,+}$.

**2.** Condition (xv) of $\mathcal{F}$ only needs to hold with exponent $2 + \delta$ for Lemma AN3(a) to hold. For Lemma AN3(b), which gives consistency of the estimator of the covariance kernel, the exponent $4 + \delta$ is needed to control the variance of the covariance estimator.

**3.** The proof of Lemma AN3(a) is an extension of the proof of Lemma A1 of

AS2 (which is given in Appendix E of AS2). The proof of Lemma AN3(b) is different from that of Lemma A1 of AS2 because the summands $m(W_i, \theta, g, b)$ are not uniformly integrable, so a standard uniform law of large numbers cannot be employed. Rather, an empirical process maximal inequality is utilized.

**4.** To prove Theorem AN1, we adjust the proof of Theorem 1 of AS1. The proof of Theorem 1 of AS1 uses a subsequence argument to reduce a uniform result over $(\theta, F) \in \mathcal{F}$ for which $h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}$ as $n \to \infty$ to a result for a subsequence $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F} : n \geq 1\}$ for which the covariance kernels $\{h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) : n \geq 1\}$ satisfy $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ for some limit $h_{2,0} \in \mathcal{H}_2$.

In AS1 and AS2, the covariance kernel $h_{2,F}(\theta)$ of $\nu_n(\theta, \cdot)$ is a normalized sum of terms $m(W_i, \theta, g)$ and does not depend on $n$. Hence, the sample-size $n$ and the asymptotic covariance kernels are the same. In contrast, in this paper, the covariance kernel $h_{2,F}(\theta, b)$ of $\nu_{n,F}(\theta, \cdot, b)$ is a normalized sum of terms $m(W_i, \theta, g, b)$ and it depends on $n$ through $b$. Here, the subsequence of covariance kernels $\{h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) : n \geq 1\}$ (that arises from the subsequence argument in AS2) is a subsequence of asymptotic kernels. We use Lemma AN1(c) to show that if $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$, then the sample-size $a_n$ covariance kernel $h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n})$ satisfies $d(h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$. This holds because

$$\begin{aligned}
& d(h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n}), h_{2,0}) \\
\leq\ & d(h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n}), h_{2,F_{a_n}}(\theta_{a_n})) + d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \\
\leq\ & \sup_{(\theta,F)\in\mathcal{F}} d(h_{2,F}(\theta, b_{a_n}), h_{2,F}(\theta)) + d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \\
\to\ & 0,
\end{aligned} \tag{9.7}$$

where the first inequality holds by the triangle inequality and the convergence holds by Lemma AN1(c). The convergence result in (9.7) is the condition that is needed to obtain the weak convergence of the empirical process $\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot, b_{a_n})$ in Lemma AN3(a).

### 9.4.4 Proofs of Theorems AN1 and N1

**Proof of Theorem AN1.** We adjust the proof of Theorem 1 of AS1 to prove Theorem AN1. The proof of Theorem 1 of AS1 is given in AS2. It goes through as stated using Lemma AN3 in place of Lemma A1 of AS2 except for one inequality. The second inequality in (12.14) of AS1 does not necessarily hold because it relies on

$h_{1,a_n,F_{a_n}}(\theta_{a_n}, g) \geq 0_p$, which does not necessarily hold when the summands $m(W_i, \theta, g)$ are replaced by $m(W_i, \theta, g, b)$.

The reason is as follows. In this paper, $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g)$ of AS2 is replaced by $h_{1,a_n,F_{a_n}}$ $(\theta_{a_n}, g, b_{a_n})$. For arbitrary $(\theta, F)$, the latter can be written as

$$h_{1,n,F}(\theta, g, b) = n^{1/2} D_F^{-1/2}(\theta, b) E_F[b^{-d_z/2} K_b(Z_i) E_F(m(W_i, \theta, g)|Z_i)] \tag{9.8}$$

using its definition in (9.2) and iterated expectations. By the conditional moment inequalities in (2.1) (or conditions (iii) and (iv) of $\mathcal{F}$ in (2.15)), $E_F(m(W_i, \theta, g)|Z_i = z) \geq 0$ when $z = z_0$. But, for other values of $z$, this inequality need not hold. In (9.8), $E_F(m(W_i, \theta, g)|Z_i = z)$ receives a non-zero weight for all $Z_i = z$ such that $K_b(z) \neq 0$. Hence, $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n}) \geq 0_p$ need not hold.

By Lemma AN2, we have $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n}) \geq -\eta_1$ for $n$ sufficiently large for some $\eta_1 > 0$. Hence, the second through fourth inequalities in (12.14) of AS1 are valid in the present context when $\tilde{\nu}_{a_n}(g)(\omega)$ and $-B_\chi(\omega)$ are replaced by $\tilde{\nu}_{a_n}(g)(\omega) - \eta_1$ and $-B_\chi(\omega) - \eta_1$, respectively. $\square$

**Proof of Theorem N1.** We adjust the proof of Theorem 2 in AS1 to prove Theorem N1. The proof of Theorem 2 of AS1 is given by the combination of Lemmas A2-A5 of AS2. Hence, we need to establish analogues of these Lemmas that hold in the context of this paper.

In the analogue of Lemma A2, the quantity $c_0(h_{n,F}(\theta), 1-\alpha)$ is replaced by $c_0(h_{n,F}(\theta, b), 1-\alpha)$ because the latter is the $1-\alpha$ quantile of the distribution of $T_n(\theta)$, which depends on $h_{n,F}(\theta, b)$, not $h_{n,F}(\theta)$. Given this change, the proof of Lemma A2 of AS2 goes through making use of Theorem AN1 in place of Theorem 1 of AS1. Note that the quantity $x_{h_{n,F}(\theta)}$ that appears in Theorem 1 and in the proof of Lemma A2 of AS2 is changed to $x_{h_{n,F}(\theta,b)}$ in Theorem AN1 because we take $x_{h_{n,F}(\theta,b)} = c_0(h_{n,F}(\theta, b), 1-\alpha) + \delta$ in the proof of the analogue of Lemma A2.

In the analogue of Lemma A3 of AS2, we use the property of the sequence $\{h_{a_n,F_{a_n}}(\theta_{a_n}): n \geq 1\}$ constructed there (that $d(h_{a_n,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$) and Lemma AN1(c) to show that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$. The rest of the proof of Lemma A3 of AS2 goes through (with the empirical process and other finite-sample quantities depending on $b_{a_n}$) except for the second inequality in (12.23) of AS2. The latter does not hold because $h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n}) \geq 0$ does not necessarily hold, as discussed in the proof of Theorem AN1 above, and hence "$\varphi_{a_n,j}(\theta_{a_n}, g) = 0 \leq h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g)$

whenever $\xi_{a_n,j}(\theta_{a_n}, g) \leq 1$" does not necessarily hold. We replace the latter with "$\varphi_{a_n,j}(\theta_{a_n}, g) = -\eta \leq h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n})$ whenever $\xi_{a_n,j}(\theta_{a_n}, g) \leq 1$ for $n$ sufficiently large," which holds by the definition of $\varphi_n(\theta, g)$ in (4.11) and Lemma AN2.

In the proof of the analogue of Lemma A4 of AS2, we use Lemma AN1(c) to show that the sequence $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\}$ constructed there is in $SubSeq(h_{2,0})$ (as in the proof of the analogue of Lemma A3). The rest of the proof of the analogue of Lemma A4 goes through with the only changes being that $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g)$ and $h_{2,F_{a_n}}(\theta_{a_n}, g)$ depend on $b_{a_n}$.

The proof of the analogue of Lemma A5 of AS2 goes through with the only change being a change in the set $\mathcal{H}_1$, which is defined in Section 5.2 of AS1 to be the set of all functions from $\mathcal{G}$ to $R^p_{+,\infty} \times \{0\}^v$, to the set $\mathcal{H}_{1,\eta_1}$, which we define here to be the set of all functions from $\mathcal{G}$ to $[-\eta_1, \infty]^p \times \{0\}^v$ for some $\eta_1 > 0$. The latter definition allows the functions in $\mathcal{H}_{1,\eta_1}$ to take small negative values, which accommodates the fact that $h_{1,n,F}(\theta, g, b)$ can be negative.

Given the analogues of Lemmas A1-A5 of AS2, the proof of Theorem N1 is complete. $\square$

#### 9.4.5 Lemmas AN4-AN6 and Proofs of Lemmas AN1-AN3

The proof of Lemma AN1 uses the following three Lemmas.

Let $A \odot B$ denote the direct (i.e., element-by-element) product of two matrices $A$ and $B$ with the same dimensions.

**Lemma AN4.** *Suppose Assumption* NM *holds. Then, for all $g, g^* \in \mathcal{G}$ and $(\theta, F) \in \mathcal{F}_+$,*

$$\Sigma_F(\theta, g, g^*, z_0) = E_F \Sigma_F(\theta, X_i, z_0) \odot (g(X_i) g^*(X_i)'),$$

*where $\Sigma_F(\theta, x, z)$ and $\Sigma_F(\theta, g, g^*, z)$ are defined in* (2.14) *and* (4.5), *respectively.*

**Lemma AN5.** *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$\sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g\in\mathcal{G}} ||b^{-d_z/2} E_F K_b(Z_i) m(W_i, \theta, g)|| = O(b^{d_z/2}) = o(1).$$

**Lemma AN6.** *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$\sup_{(\theta,F)\in\mathcal{F}_+}\sup_{g,g^*\in\mathcal{G}}||b^{-d_z}E_FK_b^2(Z_i)m(W_i,\theta,g)m(W_i,\theta,g^*)'$$
$$-E_F\Sigma_F(\theta,X_i,z_0)\odot(g(X_i)g^*(X_i)')||\to 0.$$

**Proof of Lemma AN1.** Using the definitions in (4.5) and (9.1), part (a) is established as follows. We have

$$\begin{aligned}
\Sigma_F(\theta,g,g^*,b) &= Cov_F(b^{-d_z/2}K_b(Z_i)m(W_i,\theta,g),b^{-d_z/2}K_b(Z_i)m(W_i,\theta,g^*))\\
&= b^{-d_z}E_FK_b^2(Z_i)m(W_i,\theta,g)m(W_i,\theta,g^*)'\\
&\quad -b^{-d_z/2}E_FK_b(Z_i)m(W_i,\theta,g)\cdot b^{-d_z/2}E_FK_b(Z_i)m(W_i,\theta,g^*)'\\
&= E_F[\Sigma_F(\theta,X_i,z_0)\odot(g(X_i)g^*(X_i)')]+o(1)\\
&= \Sigma_F(\theta,g,g^*,z_0)+o(1), \qquad (9.9)
\end{aligned}$$

where the $o(1)$ term holds uniformly over $g,g^*\in\mathcal{G}$ and $(\theta,F)\in\mathcal{F}_+$, the third equality holds by Lemmas AN5 and AN6, and the fourth equality holds by Lemma AN4.

Part (b) follows from part (a) by taking $g=g^*=1_k$ because $D_F(\theta,b)=Diag(\Sigma_F(\theta,1_k,1_k,b))$, $D_F(\theta,z_0)=Diag(\Sigma_F(\theta,1_k,1_k,z_0))$, and $\sup_{(\theta,F)\in\mathcal{F}_+}\left\|D_F^{-1}(\theta,z_0)\right\|<\infty$ by condition (x) of $\mathcal{F}$ in (2.15).

Part (c) follows from parts (a) and (b) because

$$\begin{aligned}
h_{2,F}(\theta,g,g^*,b) &= \left[D_F^{-1/2}(\theta,b)D_F^{1/2}(\theta,z_0)\right]\left[D_F^{-1/2}(\theta,z_0)\Sigma_F(\theta,g,g^*,b)D_F^{-1/2}(\theta,z_0)\right]\\
&\quad\times\left[D_F^{1/2}(\theta,z_0)D_F^{-1/2}(\theta,b)\right],\\
h_{2,F}(\theta,g,g^*,z_0) &= D_F^{-1/2}(\theta,z_0)\Sigma_F(\theta,g,g^*,z_0)D_F^{-1/2}(\theta,z_0), \qquad (9.10)
\end{aligned}$$

and $\sup_{(\theta,F)\in\mathcal{F}_+}\left\|D_F^{-1/2}(\theta,z_0)\right\|<\infty$. $\square$

**Proof of Lemma AN2.** For notational simplicity, suppose $m_F(\theta,x,z)$ is a scalar. This is without loss of generality (wlog) because we could argue element by element. By a two-term Taylor expansion of $m_F(\theta, x, z_0 + bz^*)$ around $z^* = 0$, we have

$$
\begin{aligned}
&\sup_{(\theta,F)\in\mathcal{F}_+} \left| \int K(z^*)\left[m_F(\theta,x,z_0+bz^*) - m_F(\theta,x,z_0)\right]dz^* \right| \\
&= \sup_{(\theta,F)\in\mathcal{F}_+} \left| b\int z^{*\prime}K(z^*)\,dz^* \frac{\partial}{\partial z} m_F(\theta,x,z_0) + \frac{b^2}{2}\int K(z^*)z^{*\prime}\frac{\partial^2}{\partial z\partial z'} m_F(\theta,x,\widetilde{z})z^* dz^* \right| \\
&\le b^2 \sup_{z\in[-1,1]^{d_z}} |K(z)| \cdot \sup_{(\theta,F)\in\mathcal{F}} \sup_{z\in\mathcal{Z}_0} \left\| \frac{\partial^2}{\partial z\partial z'} m_F(\theta,x,z) \right\| \cdot \left| \int_{[-1,1]^{d_z}} z^{*\prime}z^* dz^* \right| \\
&= b^2 L_m(x) C
\end{aligned} \tag{9.11}
$$

for some $C < \infty$, where the Taylor expansion is valid by condition (xi) of $\mathcal{F}$ in (2.15), $\widetilde{z}$ is some intermediate point that is in $\mathcal{Z}_0$ for $b$ sufficiently small, the inequality uses Assumption K(c), the last equality uses Assumptions K(d) and K(e), and $L_m(x)$ is defined in condition (xi) of $\mathcal{F}$ in (2.15).

Using (9.11), we have: for all $(\theta, F) \in \mathcal{F}$ and $g \in \mathcal{G}$,

$$
\begin{aligned}
&|E_F m(W_i,\theta,g,b) - b^{d_z/2}E_F m_F(\theta,X_i,z_0)g(X_i)| \\
&= |b^{-d_z/2}E_F K_b(Z_i)m(W_i,\theta)g(X_i) - b^{d_z/2}E_F m_F(\theta,X_i,z_0)g(X_i)| \\
&= \left| \int \left( \int b^{-d_z/2} K\left(\frac{z-z_0}{b}\right) m_F(\theta,x,z)dz - b^{d_z/2} m_F(\theta,x,z_0) \right) g(x)f(x)d\mu_X(x) \right| \\
&= b^{d_z/2} \left| \int \left( \int \left[K(z^*)\, m_F(\theta,x,z_0+bz^*) - K(z^*)\, m_F(\theta,x,z_0)\right] dz^* \right) g(x)f(x)d\mu_X(x) \right| \\
&\le b^{d_z/2} \int b^2 L_m(x) C G f(x) d\mu_X(x) \\
&\le b^{2+d_z/2} C G C_2,
\end{aligned} \tag{9.12}
$$

where $CGC_2 < \infty$, the first equality holds by the definition of $m(W_i, \theta, g, b)$, the second equality uses iterated expectations with conditioning on $(X_i, Z_i)$ and the definition of $m_F(\theta, x, z)$, the third equality holds by change of variables with $z^* = (z - z_0)/b$, the first inequality holds by (9.11) and Assumption NM(a), and the second inequality holds by condition (xi) of $\mathcal{F}$ in (2.15).

By Assumption B(a), $n^{1/2}O(b^{2+d_z/2}) = o(1)$. This and (9.12) give

$$
\sup_{(\theta,F)\in\mathcal{F}} \sup_{g\in\mathcal{G}} |n^{1/2}E_F m(W_i,\theta,g,b) - (nb^{d_z})^{1/2} E_F m_F(\theta,X_i,z_0)g(X_i)| = o(1). \tag{9.13}
$$

Equations (9.11)-(9.13) also hold with $D_F^{-1/2}(\theta, z_0)$ multiplying each quantity inside the absolute values using condition (x) of $\mathcal{F}$ in (2.15). Equation (9.13) (with the multiplicand $D_F^{-1/2}(\theta, z_0)$ added inside the absolute values), Lemma AN1(b), and the definition of $h_{1,n,F}(\theta, g, b)$ give

$$\sup_{(\theta,F)\in\mathcal{F}} \sup_{g\in\mathcal{G}} |(I_k + o(1))h_{1,n,F}(\theta, g, b) - (nb^{d_z})^{1/2} D_F^{-1/2}(\theta, z_0) E_F m_F(\theta, X_i, z_0) g(X_i)| = o(1). \tag{9.14}$$

By conditions (iii) and (x) of $\mathcal{F}$ in (2.15), $m_F(\theta, x, z_0) \geq 0$ a.s. $[F_X]$ $\forall(\theta, F) \in \mathcal{F}$ and $D_F^{-1/2}(\theta, z_0)$ is pd. In addition, $g(x) \geq 0$ $\forall x \in R^{d_x}$ by Assumption NM(a). Hence, for all $(\theta, F) \in \mathcal{F}$ and $g \in \mathcal{G}$,

$$(nb^{d_z})^{1/2} D_F^{-1/2}(\theta, z_0) E_F m_F(\theta, X_i, z_0) g(X_i) \geq 0. \tag{9.15}$$

Equations (9.14) and (9.15) combine to establish the result of the Lemma. $\square$

**Proof of Lemma AN3.** The proof of part (a) follows the same argument as used to prove Lemma A1(a) of AS2 using Lemmas E1-E3 in Appendix E of AS2. Lemmas E1 and E2 hold without change.

The results of Lemma E3 of AS2 hold for $SubSeq(h_2)$ as defined here with $h_2 \in \mathcal{H}_{2,+}$ and with $m(W_{n,i}(\omega), \theta_n, g)$ and $D_{F_n}^{-1/2}(\theta_n)$ replaced by $m(W_{n,i}(\omega), \theta_n, g, b)$ and $D_{F_n}^{-1/2}(\theta_n, b)$, respectively, in (16.4) of AS2. Lemma E3 of AS2 is proved by verifying conditions (i)-(v) of Theorem 10.6 of Pollard (1990). The proof in the present context requires some adjustments.

In the verification of (i), $m(W_{n,i}(\omega), \theta_n, g)$ and $\sigma_{F_n,j}(\theta_n)$ are replaced by $m(W_{n,i}(\omega), \theta_n, g, b)$ and the $(j, j)$ element of $D_{F_n}^{1/2}(\theta_n, b)$ in (16.35)-(16.36) of AS2.

In the verification of (ii), $D_{F_n}(\theta_n)$ and $\Sigma_{F_n}(\theta_n, g, g^*)$ are replaced by $D_{F_n}(\theta_n, b)$ and $\Sigma_{F_n}(\theta_n, g, g^*, b)$ in (16.37) of AS2. Then, condition (i) of $SubSeq(h_2)$ plus Lemma AN1(c) deliver the desired convergence. Lemma AN1(c) is required in the proof in the current case, but not in AS2, because the finite-sample covariance kernel of the empirical process depends on $b$ in the present case.

In the verification of (iii), one can ignore the $\sigma_{F_n,j}^{-1}(\theta_n)$ and $G(X_i)$ multiplicands in (16.38) of AS2 because Lemma AN1(b) and condition (x) of $\mathcal{F}$ in (2.15) imply that $\sigma_{F_n,j}^{-1}(\theta_n)$ is uniformly bounded over $(\theta, F) \in \mathcal{F}_+$ and $n \geq 1$ and Assumption NM(a) implies that $G(X_i) = G < \infty$. Then, Lemma AN1(a) gives the desired result.

Condition (iv) is the Lindeberg condition. In the verification of (iv), one can ignore the $\sigma_{F_n,j}^{-1}(\theta_n)$ and $G(X_i)$ multiplicands in (16.39) of AS2 for the same reasons as above. The required condition reduces to: for all $\xi > 0$, some $\delta > 0$, and all $j \leq k$,

$$
\begin{aligned}
& A_n = \sum_{i=1}^{n} E_{F_n} m_j^2(W_i, \theta_n, b) 1(|m_{n,j}(W_i, \theta_n, b)| > \xi) \to 0, \text{ where} \\
& m_{n,j}(W_i, \theta, b) = n^{-1/2} b^{-d_z/2} K_b(Z_i) m_j(W_i, \theta). \qquad (9.16)
\end{aligned}
$$

We have

$$
\begin{aligned}
A_n &\leq nE_{F_n}|m_{n,j}(W_i, \theta_n, b)|^{2+\delta}/\xi^{2+\delta} \\
&= n^{-\delta/2} b^{-\delta d_z/2} \left(b^{-d_z} E_{F_n} K_b^{2+\delta}(Z_i)|m_j(W_i, \theta_n)|^{2+\delta}/\xi^{2+\delta}\right) \\
&= (nb^{d_z})^{-\delta/2} \left(b^{-d_z} \int K^{2+\delta}\left(\frac{z - z_0}{b}\right) E_{F_n}(|m_j(W_i, \theta_n)|^{2+\delta}|Z_i = z) f_n(z) dz/\xi^{2+\delta}\right) \\
&= (nb^{d_z})^{-\delta/2} \left(\int K^{2+\delta}(z^*) E_{F_n}(|m_j(W_i, \theta_n)|^{2+\delta}|Z_i = z_0 + bz^*) f_n(z_0 + bz^*) dz^*/\xi^{2+\delta}\right) \\
&\leq (nb^{d_z})^{-\delta/2} \left(C_5^* \int K^{2+\delta}(z^*) dz^*/\xi^{2+\delta}\right) \\
&\to 0 \qquad (9.17)
\end{aligned}
$$

for some constant $C_5^* < \infty$, where the first inequality holds using identical distributions, the first equality holds by algebra, the second equality holds by iterated expectations, the third equality holds by change of variables with $z^* = (z - z_0)/b$, the second inequality holds for $b$ sufficiently small that $z_0 + bz^* \in \mathcal{Z}_0$ by condition (xiv) of $\mathcal{F}$ in (2.15), and the convergence holds by Assumptions B(b), K(c), and K(e).

In the verification of (v), $D_{F_n}(\theta_n)$ and $m(W_i, \theta_n, g)$ are replaced by $D_{F_n}(\theta_n, b)$ and $m(W_i, \theta_n, g, b)$ in (16.40) of Section 16.6 in Appendix E of AS2 and the convergence holds by condition (i) of $SubSeq(h_2)$ plus Lemma AN1(c). This completes the changes needed in the proof of Lemma E3 of AS2.

Given that the results of Lemma E3 of AS2 hold for $SubSeq(h_2)$ as defined here, the proof of Lemma A1(a) in AS2 establishes Lemma AN3(a) with only minor changes. In particular, $D_{F_n}(\theta_n)$ is replaced by $D_{F_n}(\theta_n, b)$ in (16.8) of AS2 and the second and last equalities in (16.8) of AS2 hold by (16.40) of AS2 with the changes described in the previous paragraph. This completes the proof of part (a) of Lemma AN3.

Now, we prove part (b) of the Lemma. The multiplicand $D_F^{-1/2}(\theta, b)$, which appears in $\widehat{h}_{2,n,F}(\theta, g, g^*, b)$, equals $D_F^{-1/2}(\theta, z_0) + o(1)$ uniformly over $(\theta, F) \in \mathcal{F}$ by Lemma

AN1(b) and $\sup_{(\theta,F)\in\mathcal{F}}||D_F^{-1/2}(\theta,z_0)||<\infty$ by condition (x) of $\mathcal{F}$ in (2.15). Hence, one can ignore the $D_F^{-1/2}(\theta,b)$ multiplicand when verifying part (b) of the Lemma. Doing so transforms $\widehat{h}_{2,n,F}(\theta,g,g^*,b)$ into $\widehat{\Sigma}_n(\theta,g,g^*)$.

Part of the proof of part (b) is similar to the proof of Lemma A1(b) of AS2. As in AS2, for notational simplicity, we establish results for the sequence $\{n\}$, rather than the subsequence $\{a_n : n\geq 1\}$. Two terms appear in the rhs of (16.16) of AS2. The second term can be shown to be $o_p(1)$. The argument is as follows. The second term (ignoring the $D_F^{-1/2}(\theta,b)$ multiplicand) is the following quantity multiplied by its transpose:

$$n^{-1}\sum_{i=1}^n m(W_i,\theta,g,b)=n^{-1}\sum_{i=1}^n b^{-d_z/2}K_b(Z_i)m_j(W_i,\theta_n)g(X_i). \tag{9.18}$$

This quantity has mean that is $o_p(1)$ by Lemma AN5. The difference between this quantity and its mean is $o_p(1)$ by Lemma E2 of AS2. The conditions of Lemma E2 are verified by the argument given in (16.18)-(16.22) of AS2 with (16.21), which verifies an $L^{1+\eta}$-boundedness condition, replaced by $L^2$-boundedness of $b^{-d_z/2}K_b(Z_i)m_j(W_i,\theta_n)g(X_i)$, which holds by Lemma AN6.

The first term appearing in (16.16) of AS2 (ignoring the $D_F^{-1/2}(\theta,b)$ multiplicand) is

$$Q_n(g,g^*)=n^{-1}\sum_{i=1}^n m(W_i,\theta,g,b)m(W_i,\theta,g^*,b)'. \tag{9.19}$$

To complete the proof of part (b), we need to show that the supremum over $(g,g^*)\in\mathcal{G}^2$ of $Q_n(g,g^*)$ minus its expectation is $o_p(1)$ under $\{(\theta_n,F_n):n\geq 1\}$. This cannot be done using the uniform law of large numbers given in Lemma E2 of AS2, as is done in the proof of Lemma A1(b) in AS2, because the summands do not satisfy an $L^{1+\eta}$-boundedness condition when $m(W_i,\theta,g)$ is replaced by $m(W_i,\theta,g,b)$.

In fact, the summands of $Q_n(g,g^*)$ do not even satisfy a uniform integrability condition, as the following calculations show. For simplicity, suppose $m(W_i,\theta)$ is a scalar and is independent of $Z_i$. Let $m_{n,i}(b)$ and $m_{n,i}$ denote $m(W_i,\theta_n,g,b)$ and $m(W_i,\theta_n,g)$, respectively. We have: for $L<\infty$,

$$\begin{aligned}
&E_{F_n}m_{n,i}^2(b)1(m_{n,i}^2(b)>L)\\
=\;&E_{F_n}b^{-d_z}K_b^2(Z_i)m_{n,i}^2 1(b^{-d_z}K_b^2(Z_i)m_{n,i}^2>L)\\
=\;&E_{F_n}\cdot E_{F_n}(b^{-d_z}K_b^2(Z_i)m_{n,i}^2 1(b^{-d_z}K_b^2(Z_i)m_{n,i}^2>L)|Z_i)
\end{aligned}$$

$$= \int b^{-d_z} K^2 \left( \frac{z - z_0}{b} \right) E_{F_n} \left( m_{n,i}^2 1 \left( b^{-d_z} K^2 \left( \frac{z - z_0}{b} \right) m_{n,i}^2 > L \right) |Z_i = z \right) f_n(z) dz$$
$$= \int K^2 (z^*) E_{F_n} (m_{n,i}^2 1(K^2(z^*) m_{n,i}^2 > L b^{d_z}) | Z_i = z_0 + b z^*) f_n(z_0 + b z^*) dz^*, \qquad (9.20)$$

where the second equality holds by iterated expectations and the fourth equality holds by change of variables with $z^* = (z - z_0)/b$. The $\limsup_{n \to \infty}$ of the rhs in (9.20) is not small for $L$ large because $b^{d_z} \to 0$. Hence, uniform integrability fails.

Instead, we show that

$$\sup_{g, g^* \in \mathcal{G}} |Q_n(g, g^*) - E_{F_n} Q_n(g, g^*)| \to_p 0 \qquad (9.21)$$

under $\{(\theta_n, F_n) : n \geq 1\}$ by using the maximal inequality (7.10) of Pollard (1990, p. 38) for manageable processes, which is applicable by Assumption NM(b) and Lemma E1 of AS2. For notational simplicity, suppose $m(W_i, \theta, g, b)$ is a scalar. (This is wlog because we can argue element by element.) The maximal inequality says that

$$E_{F_n} \sup_{g, g^* \in \mathcal{G}} |Q_n(g, g^*) - E_{F_n} Q_n(g, g^*)| \leq n^{-1} C E_{F_n} ||F_n^*|| \leq n C (E_{F_n} ||F_n^*||^2)^{1/2}, \qquad (9.22)$$

where $C$ is some finite constant and $F_n^*$ (using Pollard's notation) is an $n$-vector of envelope functions that satisfies $F_n^* = (F_{n,1}^*, ..., F_{n,n}^*)'$, $||F_n^*||^2 = \sum_{i=1}^n F_{n,i}^{*2}$, and

$$F_{n,i}^* = b^{-d_z} K_b^2(Z_i) ||m(W_i, \theta_n)||^2 G^2 \geq \sup_{g, g^* \in \mathcal{G}} ||m(W_i, \theta_n, g, b) m(W_i, \theta_n, g^*, b)||. \qquad (9.23)$$

We have

$$\begin{aligned}
& n^{-1} (E_{F_n} ||F_n^*||^2)^{1/2} \\
&= n^{-1/2} (E_{F_n} F_{n,1}^{*2})^{1/2} \\
&= n^{-1/2} G^2 (E_{F_n} b^{-2d_z} K_b^4(Z_i) ||m(W_i, \theta_n)||^4)^{1/2} \\
&= (n b^{d_z})^{-1/2} G^2 \left( \int b^{-d_z} K^4 \left( \frac{z - z_0}{b} \right) E_{F_n} (||m(W_i, \theta_n)||^4 | Z_i = z) f_n(z) dz \right)^{1/2} \\
&= (n b^{d_z})^{-1/2} G^2 \left( \int K^4(z^*) E_{F_n} (||m(W_i, \theta_n)||^4 | Z_i = z_0 + b z^*) f_n(z_0 + b z^*) dz^* \right)^{1/2}
\end{aligned}$$

$$\leq (nb^{d_z})^{-1/2}G^2\left(\int K^4(z^*)dz^* \sup_{(\theta,F)\in\mathcal{F}_+}\sup_{z\in\mathcal{Z}_0} E_F(||m(W_i,\theta)||^4|Z_i=z)f(z)\right)^{1/2}$$
$$\to 0, \tag{9.24}$$

where the first equality holds by identical distributions for $i = 1, ..., n$ under $F_n$, the second equality holds using Assumption NM(a), the third equality holds by iterated expectations, the fourth equality holds by change of variables with $z^* = (z - z_0)/b$, the inequality holds for $b$ sufficiently small using Assumption K(c), and the convergence holds by Assumptions B(b) and K(c)-(e) and condition (xiv) of $\mathcal{F}$ in (2.15). This completes the proof of part (b) of the Lemma. □

### 9.4.6 Proofs of Lemmas AN4-AN6

**Proof of Lemma AN4.** Using conditions (v)-(viii) of $\mathcal{F}$ in (2.15) (which also hold for $\mathcal{F}_+$), we have

$$\begin{aligned}\Sigma_F(\theta, g, g^*, z) &= E_F(m(W_i,\theta,g)m(W_i,\theta,g^*)'|Z_i = z)f(z)\\ &= \int\int m(y,x,z,\theta,g)m(y,x,z,\theta,g^*)'f(y,x|z)d\mu_Y(y)d\mu_X(x)f(z)\\ &= \int\int m(y,x,z,\theta,g)m(y,x,z,\theta,g^*)'f(y,x,z)d\mu_Y(y)d\mu_X(x).\end{aligned} \tag{9.25}$$

In addition, we have

$$\begin{aligned}&E_F[\Sigma_F(\theta, X_i, z)\odot(g(X_i)g^*(X_i)')]\\ &= \int[\Sigma_F(\theta,x,z)\odot(g(x)g^*(x)')]f(x)d\mu_X(x)\\ &= \int\left[\int m(y,x,z,\theta)m(y,x,z,\theta)'f(y|x,z)d\mu_Y(y)f(z|x)\odot(g(x)g^*(x)')\right]f(x)d\mu_X(x)\\ &= \int\int m(y,x,z,\theta,g)m(y,x,z,\theta,g^*)'f(y,x,z)d\mu_Y(y)d\mu_X(x),\end{aligned} \tag{9.26}$$

where the last equality uses $m(w,\theta,g) = m(w,\theta)\odot g(x)$ for $w = (y,x,z)'$. □

**Proof of Lemma AN5.** Define

$$m_F(\theta, g, z) = E_F(m(W_i,\theta,g)|Z_i = z)f(z). \tag{9.27}$$

We have

$$\begin{aligned}
&\sup_{(\theta,F)\in\mathcal{F}_+}\sup_{g\in\mathcal{G}}||b^{-d_z/2}E_F K_b(Z_i)m(W_i,\theta,g)|| \\
&= \sup_{(\theta,F)\in\mathcal{F}_+}\sup_{g\in\mathcal{G}}||b^{-d_z/2}\int K_b(z)m_F(\theta,g,z)dz|| \\
&\leq b^{-d_z/2}\int K\left(\frac{z-z_0}{b}\right)\sup_{(\theta,F)\in\mathcal{F}_+}\sup_{g\in\mathcal{G}}||m_F(\theta,g,z)||dz \\
&= b^{d_z/2}\int K(z^*)\sup_{(\theta,F)\in\mathcal{F}_+}\sup_{g\in\mathcal{G}}||m_F(\theta,g,z_0+bz^*)||dz^* \\
&\leq b^{d_z/2}\sup_{(\theta,F)\in\mathcal{F}_+}\sup_{z\in\mathcal{Z}_0}\sup_{g\in\mathcal{G}}||m_F(\theta,g,z)|| \\
&\to 0, \qquad (9.28)
\end{aligned}$$

where the first equality holds by iterated expectations conditioning on $Z_i$ using condition (v) of $\mathcal{F}$ in (2.15), the second equality holds by change of variables with $z^* = (z - z_0)/b$, the second inequality holds using Assumption K(a), and the convergence holds by Assumption B(a) and the result:

$$\sup_{(\theta,F)\in\mathcal{F}_+,z\in\mathcal{Z}_0,g\in\mathcal{G}}||m_F(\theta,g,z)|| < \infty. \qquad (9.29)$$

Equation (9.29) is established as follows. We have

$$\begin{aligned}
m_F(\theta,g,z) &= E_F E_F(m(W_i,\theta,g)|X_i,Z_i=z)f(z) \\
&= \int E_F(m(W_i,\theta)|X_i=x,Z_i=z)g(x)f(x|z)d\mu_X(x)f(z) \\
&= \int m_F(\theta,x,z)g(x)f(x,z)d\mu_X(x), \qquad (9.30)
\end{aligned}$$

where the second equality uses condition (ix) of $\mathcal{F}$ in (2.15). Hence, we obtain

$$\begin{aligned}
&\sup_{(\theta,F)\in\mathcal{F}_+}\sup_{z\in\mathcal{Z}_0}\sup_{g\in\mathcal{G}}||m_F(\theta,g,z)|| \\
&\leq G\sup_{(\theta,F)\in\mathcal{F}_+}\sup_{z\in\mathcal{Z}_0}\int||m_F(\theta,x,z)||f(x,z)d\mu_X(x) < \infty, \qquad (9.31)
\end{aligned}$$

where the first inequality holds by Assumption NM(a) and the second inequality holds by condition (xii) of $\mathcal{F}$ in (2.15). □

**Proof of Lemma AN6.** For notational simplicity, we suppose $m(W_i, \theta, g)$ is a scalar. (This is wlog because we could argue element by element.) For all $g, g^* \in \mathcal{G}$, we have

$$\begin{aligned}
& J_b(g, g^*) \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} |b^{-d_z} E_F K_b^2(Z_i) m^2(W_i, \theta, g) - E_F[\Sigma_F(\theta, X_i, z_0) \odot (g(X_i) g^*(X_i))]| \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} |b^{-d_z} E_F K_b^2(Z_i) m^2(W_i, \theta) g(X_i) g^*(X_i) - E_F \Sigma_F(\theta, X_i, z_0) g(X_i) g^*(X_i)| \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} \left| \int \left( \int b^{-d_z} K^2 \left( \frac{z - z_0}{b} \right) \Sigma_F(\theta, x, z) dz - \Sigma_F(\theta, x, z_0) \right) g(x) g^*(x) f(x) d\mu_X(x) \right| \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} \left| \int \left( \int \left[ K^2(z^*) \Sigma_F(\theta, x, z_0 + b z^*) - K^2(z^*) \Sigma_F(\theta, x, z_0) \right] dz^* \right) \right. \\
& \left. \times g(x) g^*(x) f(x) d\mu_X(x) \right|, \qquad (9.32)
\end{aligned}$$

where the first equality defines $J_b(g, g^*)$, the second equality holds by the definition of $m(W_i, \theta, g)$, the third equality uses iterated expectations with conditioning on $(X_i, Z_i)$ and conditions (vi) and (viii) of $\mathcal{F}$ in (2.15), and the fourth equality holds by change of variables with $z^* = (z - z_0)/b$.

Using (9.32), we have

$$\begin{aligned}
\sup_{g,g^*\in\mathcal{G}} J_b(g, g^*) & \leq G \sup_{(\theta,F)\in\mathcal{F}_+} \int \left( \int K^2(z^*) L_\Sigma(x) b ||z^*|| dz^* \right) f(x) d\mu_X(x) \\
& \leq bGC \sup_{(\theta,F)\in\mathcal{F}_+} \int L_\Sigma(x) f(x) d\mu_X(x) \\
& \to 0, \qquad (9.33)
\end{aligned}$$

where the first inequality holds by condition (xiii) of $\mathcal{F}$ in (2.15) and Assumption NM(a), the second inequality holds for some $C < \infty$ by Assumptions K(c) and K(e), and the convergence holds by Assumptions B(a) and condition (xiii) of $\mathcal{F}$ in (2.15). $\square$

# References


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., Jia, P., 2008. Inference for parameters defined by moment inequalities: a recommended moment selection procedure. Cowles Foundation Discussion Paper No. 1676, Yale University.

Andrews, D. W. K., Shi, X., 2007a. Inference based on conditional moment inequalities. Cowles Foundation Discussion Paper No. 1761R, Yale University.

Andrews, D. W. K., Shi, X., 2007b. Supplement to "inference based on conditional moment inequalities." Cowles Foundation Discussion Paper No. 1761R, Yale University.

Andrews, D. W. K., Shi, X., 2010a. Appendix 2 to "Nonparametric inference based on conditional moment inequalities." Cowles Foundation Discussion Paper No. 1840, Yale University.

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

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., Ghandi, A. Quint, D., 2010. Identification and testing in ascending auctions with unobserved heterogeneity. Unpublished manuscript, Department of Economics, University of Wisconsin, Madison.

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

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., 2010. Sharp identification regions in models with convex moment predictions. CEMMAP Working Paper CWP25/10, Institute for Fiscal Studies, UCL.

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., Rozen, A., 2008. Inference with intersection bounds. Unpublished manuscript, Department of Economics, University College London.

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.

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.

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., 2008. Estimation and inference with many moment inequalities. Unpublished manuscript, Department of Economics, MIT.

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., 2008. Misspecification in moment inequality models: back to moment equalities? Unpublished manuscript, Department of Economics, Northwestern University.

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 Mcmillan, New York.

# Appendix 2

to

# 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, Madison

June 2010
Revised: December 2011

# 10 Appendix 2

This Appendix provides proofs of Theorems N2 and N3 of the paper "Nonparametric inference based on conditional moment inequalities." It also provides some additional simulation results to those given in that paper. We let AS1 and AS2 abbreviate Andrews and Shi (2007a) and Andrews and Shi (2007b), respectively.

## 10.1 Proofs of Theorems N2 and N3

**Proof of Theorem N2.** Theorem N2 is analogous to Theorem 3 of AS1. The proof of Theorem 3 of AS1 that is given in Section 14.2 in Appendix C of AS2 goes through with a few changes in the present context. First, $E_{F_0}(\cdot)$ is replaced by $E_{F_0}(\cdot|Z_i = z_0)$ in $m^*(g)$ and elsewhere. Second, $n^{1/2}\beta(g_0)$ is replaced throughout by $(nb^{d_z})^{1/2}\beta(g_0)$. Third, Assumption NFA(a) is used in place of Assumption FA(a) to obtain the inequality in (14.28) of AS2. Fourth, the proof uses Lemma AN3, which employs Assumptions NFA(b) and NFA(c), in place of Lemma A1 of AS2.

Fifth, the second equality of (14.33) of AS2 does not hold. It relies on $n^{-1/2}h_{1,n,F_0}(\theta_*, g)$ $= m^*(g)$, which in the present context is replaced by $(nb^{d_z})^{-1/2}h_{1,n,F_0}(\theta_*, g, b) = m^*(g)$, which does not hold. However, we have

$$\begin{aligned}
(nb^{d_z})^{-1/2}h_{1,n,F_0}(\theta_*, g, b) &= D_{F_0}^{-1/2}(\theta_*, b)b^{-d_z/2}E_{F_0}m(W_i, \theta_*, g, b) \\
&= D_{F_0}^{-1/2}(\theta_*, z_0)E_{F_0}m(\theta_*, X_i, z_0)g(X_i) + O(b^2) \\
&= D_{F_0}^{-1/2}(\theta_*, z_0)E_{F_0}(m(W_i, \theta_*, g)|Z_i = z_0)f(z_0) + O(b^2) \\
&= m^*(g) + o(1), \qquad (10.34)
\end{aligned}$$

where the second equality holds by Lemma AN1(b) and (9.12) (which holds for $(\theta_*, F_0) \in \mathcal{F}_+$), the third equality holds by the same argument as in the proof of Lemma AN4 with $m(y, x, z, \theta, g)m(y, x, z, \theta, g^*)'$ replaced by $m(y, x, z, \theta, g)$ throughout, and the fourth equality holds by the definition of $m^*(g)$ and Assumption B(a).

Using (10.34), the second equality of (14.33) of AS2 holds with $m^*(g)/\beta(g_0)$ replaced by $m^*(g)/\beta(g_0) + o(1)$.

These are the only changes needed to the proof of Theorem 3 of AS1. □

**Proof of Theorem N3.** Theorem N3 is analogous to Theorem 4 of AS1. First, we give an analogue of (14.37) in the proof of Theorem 4 of AS1 given in Section 14.3 of

Appendix C in AS2. We have

$$
\begin{aligned}
& h_{1,n,F_n}(\theta_{n,*}, g, b) \\
= & \; n^{1/2} D_{F_n}^{-1/2}(\theta_{n,*}, b) E_{F_n} m(W_i, \theta_{n,*}, g, b) \\
= & \; (nb^{d_z})^{1/2}(I_k + o(1)) D_{F_n}^{-1/2}(\theta_{n,*}, z_0) E_{F_n} m(\theta_{n,*}, X_i, z_0) g(X_i) + o(1) \qquad (10.35) \\
= & \; (nb^{d_z})^{1/2}(I_k + o(1)) D_{F_n}^{-1/2}(\theta_{n,*}, z_0) E_{F_n}(m(W_i, \theta_{n,*}, g) | Z_i = z_0) f_n(z_0) + o(1),
\end{aligned}
$$

where the first equality holds by (9.2), the second equality holds by Lemma AN1(b) and (9.12) because $n^{1/2} b^{2+d_z/2} \to 0$ if $b = o(n^{-1/(4+d_z)})$, and the third equality holds by the same argument as in the proof of Lemma AN4 above.

Next, by element-by-element mean-value expansions about $\theta_n$, we have

$$
\begin{aligned}
& D_{F_n}^{-1/2}(\theta_{n,*}, z_0) E_{F_n}(m(W_i, \theta_{n,*}, g) | Z_i = z_0) f_n(z_0) \\
= & \; D_{F_n}^{-1/2}(\theta_n, z_0) E_{F_n}(m(W_i, \theta_n, g) | Z_i = z_0) f_n(z_0) \\
& + \Pi_{F_n}(\theta_{n,g}, g)(\theta_{n,*} - \theta_n), \qquad (10.36)
\end{aligned}
$$

using Assumption NLA2, where $\theta_{n,g}$ may differ across rows of $\Pi_{F_n}(\theta_{n,g}, g)$, $\theta_{n,g}$ lies between $\theta_{n,*}$ and $\theta_n$, and $\theta_{n,g} \to \theta_0$.

Combining (10.35) and (10.36) gives the analogue of (14.37) of AS2:

$$
\begin{aligned}
& h_{1,n,F_n}(\theta_{n,*}, g, b) \\
= & \; (nb^{d_z})^{1/2}(I_k + o(1)) D_{F_n}^{-1/2}(\theta_n, z_0) E_{F_n}(m(W_i, \theta_n, g) | Z_i = z_0) f_n(z_0) \\
& + (I_k + o(1)) \Pi_{F_n}(\theta_{n,g}, g)(nb^{d_z})^{1/2}(\theta_{n,*} - \theta_n) \\
\to & \; h_1(g) + \Pi_0(g) \lambda, \qquad (10.37)
\end{aligned}
$$

where $h_1(g)$ and $\Pi_0(g)$ are defined in (7.2) and the convergence uses Assumptions NLA1(a), NLA1(b), and NLA2.

Now, the proof of Theorem N3 is similar to the proof of Theorem 4 of AS1 given in AS2 with the following changes:

(i) $\{(\theta_{n,*}, F_n) \in \mathcal{F} : n \geq 1\} \in SubSeq(h_2)$, where $h_2 = h_{2,F_0}(\theta_0) \in \mathcal{H}_{2,+}$ by Assumptions NLA1(a) and NLA1(c)-(e),

(ii) part (i) and Assumptions B and MN imply that the results of Lemma AN3 hold under $\{(\theta_{n,*}, F_n) \in \mathcal{F} : n \geq 1\}$ and these results are used in place of Lemma A1 of AS2,

(iii) equation (14.38) of AS2 is replaced by

$$
\begin{aligned}
&\kappa_n^{-1}\overline{D}_{F_n}^{-1/2}(\theta_{n,*},g,b)D_{F_n}^{1/2}(\theta_{n,*},b)h_{1,n,F_n}(\theta_{n,*},g,b)\\
&= (I_k+o(1))\kappa_n^{-1}(nb^{d_z})^{1/2}\overline{D}_{F_n}^{-1/2}(\theta_n,g,z_0)E_{F_n}(m(W_i,\theta_n,g)|Z_i=z_0)f_n(z_0)\\
&\quad +\kappa_n^{-1}\overline{D}_{F_0}^{-1/2}(\theta_0,g,z_0)D_{F_0}^{1/2}(\theta_0,z_0)(I_k+o(1))\Pi_{F_n}(\theta_{n,g},g)(nb^{d_z})^{1/2}(\theta_{n,*}-\theta_n)]\\
&= \pi_1(g)+o(1), \qquad (10.38)
\end{aligned}
$$

where the first equality holds by the equality in (10.37) and Lemma AN1(b) and the second equality holds because (a) the first term on the rhs of the first equality is $\pi_1(g)+o(1)$ by Assumption NLA4 and (b) the second term on the rhs of the first equality is $o(1)$ by the convergence of the second term in (10.37) plus $\kappa_n^{-1}\to 0$, and

(iv) in the verification of (14.23) in part (ix) of the proof of Theorem 4 of AS1 given in Section 14.3 of Appendix C in AS2, (10.37) is used in place of (14.37) of AS2. This completes the proof. □

## 10.2 Additional Simulation Results

In this section, we provide some additional simulation results. Tables A1 and A2 report the robustness results for the CvM/Max and KS/Max test statistics in the kinked and the peaked bound cases, respectively, for the quantile selection model. As in Tables I-III, the results in Tables A1 and A2 are for the lower endpoints of the identified intervals. Tables A3 and A4 report the robustness results for the CvM and KS test statistics in the kinked and tilted bound cases, respectively, for the conditional treatment effect model.

Both Tables A1 and A2 show that there is little sensitivity to $r_1$, $\varepsilon$, the GMS tuning parameters, and the kernel bandwidth in terms of coverage probabilities. There is some sensitivity in terms of the FCP's. The FCP decreases (gets better) with the sample size for the KS/MAX-GMS/Asy pair and is stable for the CvM/Max-GMS/Asy pair. The FCP is smaller (better) with $(\kappa_n, B_n)$ halved and bigger with $(\kappa_n, B_n)$ doubled.

There is quite a bit sensitivity to the kernel bandwidth. With both the kinked and the peaked bound, doubling the bandwidth reduces the FCP's for tests with the KS/Max statistics. The same is true with the kinked bound and the CvM/Max statistic. However, with the peaked bound, both doubling and halving the bandwidth increases the FCP's.

Tables A1 and A2 show that 0.50 CI's cover the true value with probability noticeably higher than 0.50. This indicates that the lower boundary point of the 0.50 CI as an estimator for the lower end point of the identified set is not median unbiased, but does not have an inward bias which has been a concern in the literature.

Table A1. Nonparametric Quantile Selection Model with Kinked 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: ($n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7}$) | | .989 | .987 | .49 | .57 |
| $n = 100$ | | .988 | .991 | .48 | .59 |
| $n = 500$ | | .989 | .991 | .45 | .54 |
| $r_1 = 2$ | | .988 | .987 | .50 | .53 |
| $r_1 = 4$ | | .990 | .989 | .48 | .60 |
| $(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$ | | .991 | .987 | .49 | .55 |
| $(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$ | | .993 | .991 | .56 | .61 |
| $\varepsilon = 1/100$ | | .989 | .987 | .47 | .57 |
| $b = 0.5b^0 n^{-2/7}$ | | .986 | .987 | .69 | .77 |
| $b = 2b^0 n^{-2/7}$ | | .997 | .995 | .35 | .45 |
| $\alpha = .5$ | | .771 | .739 | .05 | .06 |
| $\alpha = .5$ & $n = 500$ | | .787 | .753 | .05 | .06 |

Table A2. Nonparametric Quantile Selection Model with Peaked 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: $(n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7})$ | | .991 | .991 | .49 | .53 |
| $n = 100$ | | .989 | .990 | .56 | .65 |
| $n = 500$ | | .994 | .995 | .50 | .45 |
| $r_1 = 2$ | | .990 | .990 | .51 | .50 |
| $r_1 = 4$ | | .992 | .991 | .48 | .58 |
| $(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$ | | .992 | .990 | .47 | .52 |
| $(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$ | | .994 | .994 | .54 | .56 |
| $\varepsilon = 1/100$ | | .991 | .991 | .47 | .53 |
| $b = 0.5b^0 n^{-2/7}$ | | .988 | .989 | .62 | .70 |
| $b = 2b^0 n^{-2/7}$ | | .997 | .996 | .53 | .47 |
| $\alpha = .5$ | | .803 | .761 | .04 | .05 |
| $\alpha = .5$ & $n = 500$ | | .836 | .795 | .04 | .04 |

Tables A3 and A4 show the sensitivity results for the nonparametric conditional treatment effect model with kinked bound and tilted bound, respectively.

Table A3 shows that, with the kinked bound, the test has NRP's smaller than 0.05 for all the test configurations and sample sizes that we experimented with. This is expected because with the kinked bound, the conditional moment inequality is only binding at a measure-zero set of the instrumental variable and Assumption GMS2 is not likely to hold. The ARP's are relatively stable as we vary $r_1$, decrease $\varepsilon$ or decrease $(\kappa_n, B_n)$. Doubling $(\kappa_n, B_n)$ makes the ARP's smaller (worse). Both doubling and halving the kernel bandwidth reduces ARP's noticeably.

Table A3. Nonparametric Conditional Treatment Effect Model with Kinked Bound: Variations on the Base Case

| Case | Statistic: / Crit Val: | (a) Null Rejection Probabilities CvM GMS/Asy | KS GMS/Asy | (b) Rej Probs under $H_1$ (NRP-corrected) CvM GMS/Asy | KS GMS/Asy |
|---|---|---|---|---|---|
| Base Case: $(n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7})$ | | .000 | .000 | .52 | .49 |
| $n = 100$ | | .000 | .000 | .65 | .55 |
| $n = 500$ | | .000 | .000 | .33 | .40 |
| $r_1 = 2$ | | .000 | .000 | .52 | .53 |
| $r_1 = 4$ | | .000 | .000 | .51 | .45 |
| $(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$ | | .000 | .000 | .52 | .52 |
| $(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$ | | .000 | .000 | .44 | .42 |
| $\varepsilon = 1/100$ | | .000 | .000 | .52 | .44 |
| $b = 0.5b^0 n^{-2/7}$ | | .000 | .000 | .38 | .30 |
| $b = 2b^0 n^{-2/7}$ | | .000 | .000 | .34 | .43 |

Table A4 shows a new aspect of the sensitivity analysis. The NRP for the CvM test in the base case is somewhat bigger than 0.05. Halving the bandwidth reduces NRP's to below 0.05. while doubling the bandwidth increases the NRP's to disastrous level. This is expected because with the tilted bound the unconditional moment formed using the kernel functions has negative expectation for any fixed bandwidth. The negative expectation converges to zero as the bandwidth converges to zero. Thus, letting $b$ converge to zero is central to the theoretical validity of our method. Using a large $b$ deviates from the asymptotic theory.

The ARP's in Table A4 are reasonably stable across different configurations and sample sizes, except that they are somewhat sensitive to the kernel bandwidth.

Table A4. Nonparametric Conditional Treatment Effect Model with Tilted Bound: Variations on the Base Case

| | | (a) Null Rejection Probabilities | | (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: $(n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7})$ | | .072 | .047 | .53 | .36 |
| $n = 100$ | | .085 | .042 | .49 | .34 |
| $n = 500$ | | .072 | .050 | .53 | .40 |
| $r_1 = 2$ | | .074 | .059 | .52 | .38 |
| $r_1 = 4$ | | .069 | .036 | .53 | .32 |
| $(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$ | | .081 | .054 | .50 | .35 |
| $(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$ | | .066 | .045 | .53 | .36 |
| $\varepsilon = 1/100$ | | .071 | .040 | .52 | .31 |
| $b = 0.5 b^0 n^{-2/7}$ | | .044 | .023 | .29 | .14 |
| $b = 2 b^0 n^{-2/7}$ | | .467 | .313 | .69 | .57 |

# References

Andrews, D. W. K., Shi, X., 2007a. Inference based on conditional moment inequalities. Cowles Foundation Discussion Paper No. 1761R, Yale University.

Andrews, D. W. K., Shi, X., 2007b. Supplement to "inference based on conditional moment inequalities." Cowles Foundation Discussion Paper No. 1761R, Yale University.