TESTING FOR A GENERAL CLASS OF FUNCTIONAL INEQUALITIES

SOKBAE LEE
Columbia University
Institute for Fiscal Studies

KYUNGCHUL SONG
University of British Columbia

YOON-JAE WHANG
Seoul National University

In this article, we propose a general method for testing inequality restrictions on nonparametric functions. Our framework includes many nonparametric testing problems in a unified framework, with a number of possible applications in auction models, game theoretic models, wage inequality, and revealed preferences. Our test involves a one-sided version of L_p functionals of kernel-type estimators ($1 \leq p < \infty$) and is easy to implement in general, mainly due to its recourse to the bootstrap method. The bootstrap procedure is based on the nonparametric bootstrap applied to kernel-based test statistics, with an option of estimating "contact sets." We provide regularity conditions under which the bootstrap test is asymptotically valid uniformly over a large class of distributions, including cases where the limiting distribution of the test statistic is degenerate. Our bootstrap test is shown to exhibit good power properties in Monte Carlo experiments, and we provide a general form of the local power function. As an illustration, we consider testing implications from auction theory, provide primitive conditions for our test, and demonstrate its usefulness by applying our test to real data. We supplement this example with the second empirical illustration in the context of wage inequality.

1. INTRODUCTION

In this article, we propose a general method for testing inequality restrictions on nonparametric functions. To describe our testing problem, let $v_{\tau,1}, \ldots, v_{\tau,J}$ denote nonparametric real-valued functions on $\mathbf{R}^d$ for each index $\tau \in \mathcal{T}$, where $\mathcal{T}$ is a

subset of a finite dimensional space. We focus on testing

$$
\begin{aligned}
&H_0 : \max\{v_{\tau,1}(x), \ldots, v_{\tau,J}(x)\} \leq 0 \text{ for all } (x, \tau) \in \mathcal{X} \times \mathcal{T}, \text{ against} \\
&H_1 : \max\{v_{\tau,1}(x), \ldots, v_{\tau,J}(x)\} > 0 \text{ for some } (x, \tau) \in \mathcal{X} \times \mathcal{T},
\end{aligned}
\tag{1.1}
$$

where $\mathcal{X} \times \mathcal{T}$ is a domain of interest. We propose a one-sided L_p integrated test statistic based on nonparametric estimators of $v_{\tau,1}, \ldots, v_{\tau,J}$. We provide general asymptotic theory for the test statistic and suggest a bootstrap procedure to compute critical values. We establish that our test has correct uniform asymptotic size and is not conservative. We also determine the asymptotic power of our test under fixed alternatives and some local alternatives.

We allow for a general class of nonparametric functions, including, as special cases, conditional mean, quantile, hazard, and distribution functions and their derivatives. For example, $v_{\tau,j}(x) = P(Y_j \leq \tau | X = x)$ can be the conditional distribution function of Y_j given $X = x$, or $v_{\tau,j}(x)$ can be the τ-th quantile of Y_j conditional on $X = x$. We also allow for transformations of these functions satisfying some regularity conditions. The nonparametric estimators we consider are mainly kernel-type estimators but can be allowed to be more general, provided that they satisfy certain Bahadur-type linear expansions.

Inequality restrictions on nonparametric functions arise often as testable implications from economic theory. For example, in first-price auctions, Guerre, Perrigne, and Vuong (2009) show that the quantiles of the observed equilibrium bid distributions with different numbers of bidders should satisfy a set of inequality restrictions (equation (5) of Guerre et al. (2009)). If the auctions are heterogeneous so that the private values are affected by observed characteristics, we may consider conditionally exogenous participation with a conditional version of the restrictions (see Sect. 3.2 of Guerre et al. (2009)). Such restrictions are in the form of multiple inequalities for linear combinations of nonparametric conditional quantile functions. Our test then can be used to test whether the restrictions hold jointly uniformly over quantiles and observed characteristics in a certain range. In this article, we use this auction example to illustrate the usefulness of our general framework. To the best of our knowledge, there does not exist an alternative test available in the literature for this kind of example.

In addition to (Guerre et al., 2009, GPV hereafter), a large number of auction models are associated with some forms of functional inequalities. See, for example, Haile and Tamer (2003), Haile, Hong, and Shum (2003), Aradillas-López, Gandhi, and Quint (2013), and Krasnokutskaya, Song, and Tang (2016) among others. Our method can be used to make inference in their setups, while allowing for continuous covariates. See Online Appendix E for further discussions on potential applications in economics, such as inference in models of games, tests of revealed preferences, and inference in partial identification.

Our framework has several distinctive merits. First, our proposal is easy to implement in general, mainly due to its recourse to the bootstrap method. The bootstrap procedure is based on nonparametric bootstrap applied to kernel-based

test statistics. We establish the general asymptotic (uniform) validity of the bootstrap procedure.

Second, our proposed test is shown to exhibit good power properties both in finite and large samples. Good power properties can be achieved by the use of critical values that adapt to the binding restrictions of functional inequalities. This could be done in various ways; in this article, we follow the "contact set" approach of Linton, Song, and Wang (2010) and propose bootstrap critical values. As is shown in this article, the bootstrap critical values yield significant power improvements. Furthermore, we find through our local power analysis that this class of tests exhibits dual convergence rates depending on Pitman directions, and in many cases, the faster of the two rates achieves a parametric rate of $\sqrt{n}$, despite the use of kernel-type test statistics.

Third, we establish the asymptotic validity of the proposed test uniformly over a large class of distributions, without imposing restrictions on the covariance structure among nonparametric estimates of $v_{\tau,j}(\cdot)$, thereby allowing for degenerate cases. Such a uniformity result is crucial for ensuring good finite sample properties for tests whose (pointwise) limiting distribution under the null hypothesis exhibits various forms of discontinuity. The discontinuity in the context of this article is highly complex, as the null hypothesis involves inequality restrictions on a multiple number of (or even a continuum of) nonparametric functions. We establish the uniform validity of the test in a way that covers these various incidences of discontinuity. Our new uniform asymptotics may be of independent interest in many other contexts.

Much of the recent literature on testing inequality restrictions focuses on conditional moment inequalities. Research on conditional moment inequalities includes Andrews and Shi (2013), Andrews and Shi (2014), Andrews and Shi (2017), Aradillas-López, Gandhi, and Quint (2016), Armstrong (2015), Armstrong and Chan (2016), Chernozhukov, Lee, and Rosen (2013), Chetverikov (2017), Fan and Park (2014), Khan and Tamer (2009), Lee, Song, and Wang (2013), and Menzel (2014) among others. In contrast, this article's approach naturally covers a wide class of inequality restrictions among nonparametric functions that the moment inequality framework does not (or at least is cumbersome to) apply. Such examples include testing multiple inequalities that are defined by differences in conditional quantile functions uniformly over covariates and quantiles. As this article demonstrates through an empirical application, such a testing problem can arise in empirical research (see Section 6).

One important class that is covered by our general approach is to test conditional stochastic dominance, which is a generalization of stochastic dominance to conditional distributions. See, for example, Chang, Lee, and Wang (2015), Delgado and Escanciano (2013) and Hsu (2017) among others. Our test can also be used as an alternative to testing monotonicity of mean regression (e.g., Chetverikov (2012) and Ghosal, Sen, and van der Vaart (2000)) as well as testing stochastic monotonicity (e.g., Lee, Linton, and Wang (2009) and Delgado and Escanciano (2012)). Our article is also related to testing affiliation in

Jun, Pinkse, and Wan (2010) and testing density ratio ordering in Beare and Moon (2015).

More generally speaking, our framework is related to testing qualitative non-parametric hypotheses such as positivity, monotonicity and convexity in non-parametric models. See, for example, Dümbgen and Spokoiny (2001), Juditsky and Nemirovski (2002) and Baraud, Huet, and Laurent (2005) among others. However, there are some important differences. First, the existing statistics literature mainly focuses on the hypothesis on the mean regression function in the ideal Gaussian white noise model, while we do not impose any parametric assumption such as Gaussianity and allow for more general functional inequalities among nonparametric functions other than the mean regression function. Second, the technical details of the existing "positivity" tests are quite different from ours.

Our current work is substantially different from our previous article, Lee et al. (2013). The latter considered only conditional moment inequalities under pointwise asymptotics. The test in Lee et al. (2013) is based on asymptotic normality under the least favorable case. The current article goes much beyond conditional moment inequalities and provides uniform validity of the bootstrap test with contact set estimation.

The remainder of the article is as follows. Section 2 gives an informal description of our general framework by introducing test statistics and critical values and by providing intuition behind our approach. In Section 3, we establish the uniform asymptotic validity of our bootstrap test. We also provide a class of distributions for which the asymptotic size is exact. In Section 4, we establish consistency of our test and its local power properties. In Section 5, we report results of some Monte Carlo experiments. In Sections 6 and 7, we give two empirical examples. The first empirical example in Section 6 is on testing auction models following GPV, and the second one in Section 7 is about testing functional inequalities via differences-in-differences in conditional quantiles, inspired by Acemoglu and Autor (2011). The empirical examples given in this section are not covered easily by existing inference methods; however, they are all special cases of our general framework. Section 8 concludes. Online Appendices provide all the proofs of theorems with a roadmap of the proofs to help readers.

2. GENERAL OVERVIEW

2.1. Test Statistics

We present a general overview of this article's framework by introducing test statistics and critical values. To ease the exposition, we confine our attention to the case of $J = 2$ here. The definitions and formal results for general J are given later in Section 3.

Throughout this article, we assume that $\mathcal{T}$ is a compact subset of a Euclidean space. a generality because when $\mathcal{T}$ is a finite set, we can redefine our test statistic by taking $\mathcal{T}$ as part of the finite index j indexing the nonparametric functions.

For $j = 1, 2$, let $\hat{v}_{\tau,j}(x)$ be a kernel-based nonparametric estimator of $v_{\tau,j}(x)$ and let its appropriately scaled version be

$$\hat{u}_{\tau,j}(x) \equiv \frac{r_{n,j}\hat{v}_{\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)},$$

where $r_{n,j}$ is an appropriate normalizing sequence that diverges to infinity,[1] and $\hat{\sigma}_{\tau,j}(x)$ is an appropriate (possibly data-dependent) scale normalization.[2] Then inference is based on the following statistic:

$$\hat{\theta} \equiv \int_{\mathcal{T}} \int_{\mathcal{X}} \max\left\{\hat{u}_{\tau,1}(x), \hat{u}_{\tau,2}(x), 0\right\}^p dx d\tau \tag{2.1}$$

$$= \int_{\mathcal{X}\times\mathcal{T}} \max\left\{\hat{u}_{\tau,1}(x), \hat{u}_{\tau,2}(x), 0\right\}^p dQ(x, \tau),$$

where Q is Lebesgue measure on $\mathcal{X} \times \mathcal{T}$. In this overview section, we focus on the case of using the max function under the integral in (2.1). Our general theory permits an alternative form $\sum_{j=1}^{2} \max\left\{\hat{u}_{\tau,j}(x), 0\right\}^p$ in place of the max function (see (3.1)).

2.2. Bootstrap Critical Values

As we shall see later, the asymptotic distribution of the test statistic exhibits complex ways of discontinuities as one perturbs the data generating processes. This suggests that the finite sample properties of the asymptotic critical values may not be stable. Furthermore, the location-scale normalization requires nonparametric estimation and thus a further choice of tuning parameters. This can worsen the finite sample properties of the critical values further. To address these issues, this article develops a bootstrap procedure.

In the following, we let $\hat{v}^*_{\tau,j}(x)$ and $\hat{\sigma}^*_{\tau,j}(x)$, $j = 1, 2$, denote the bootstrap counterparts of $\hat{v}_{\tau,j}(x)$ and $\hat{\sigma}_{\tau,j}(x)$, $j = 1, 2$. Let the bootstrap counterparts be constructed in the same way as the nonparametric estimators $\hat{v}_{\tau,j}(x)$ and $\hat{\sigma}_{\tau,j}(x)$, $j = 1, 2$, with the bootstrap sample independently drawn with replacement from the empirical distribution of the original sample. We let

$$\hat{s}^*_{\tau,j}(x) \equiv \frac{r_{n,j}\{\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x)\}}{\hat{\sigma}^*_{\tau,j}(x)}, \quad j = 1, 2. \tag{2.2}$$

Note that $\hat{s}^*_{\tau,j}(x)$ is a centered and scale normalized version of the bootstrap quantity $\hat{v}^*_{\tau,j}(x)$. Using these bootstrap quantities, we consider two versions of bootstrap critical values: one based on the least favorable case and the other based on estimating a contact set.

2.2.1. The Least Favorable Case. Under the least favorable configuration (LFC), we construct a bootstrap version of the right hand side of (2.1) as

$$\hat{\theta}^*_{\text{LFC}} \equiv \int \max\left\{\hat{s}^*_{\tau,1}(x), \hat{s}^*_{\tau,2}(x), 0\right\}^p dQ(x, \tau).$$

Under regularity conditions, bootstrap critical values based on the LFC can be shown to yield tests that are asymptotically valid *uniformly* in P. However, they are often too conservative in practice. As an alternative to the LFC-based bootstrap critical value, we propose a bootstrap critical value that can be less conservative but at the expense of introducing an additional tuning parameter.

2.2.2. Estimating a Contact Set.   As we shall show formally in a more general form in Lemma 1 in Section 3 below, it is satisfied that under H_0, for each sequence $c_n \to \infty$ such that $\sqrt{\log n}/c_n \to 0$ as $n \to \infty$,

$$
\hat{\theta} = \int_{B_{n,\{1\}}(c_n)} \max\{\hat{u}_{\tau,1}(x),0\}^P \, dQ(x,\tau)
$$
$$
+ \int_{B_{n,\{2\}}(c_n)} \max\{\hat{u}_{\tau,2}(x),0\}^P \, dQ(x,\tau)
$$
$$
+ \int_{B_{n,\{1,2\}}(c_n)} \max\{\hat{u}_{\tau,1}(x),\hat{u}_{\tau,2}(x),0\}^P \, dQ(x,\tau), \tag{2.3}
$$

with probability approaching one, where, letting $u_{n,\tau,j}(x) \equiv r_{n,j} v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)$, i.e., a population version of $\hat{u}_{\tau,j}(x)$,[3] we define

$$
B_{n,\{1\}}(c_n) \equiv \{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |u_{n,\tau,1}(x)| \leq c_n \text{ and } u_{n,\tau,2}(x) < -c_n\},
$$
$$
B_{n,\{2\}}(c_n) \equiv \{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |u_{n,\tau,2}(x)| \leq c_n \text{ and } u_{n,\tau,1}(x) < -c_n\} \text{ and}
$$
$$
B_{n,\{1,2\}}(c_n) \equiv \{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |u_{n,\tau,1}(x)| \leq c_n \text{ and } |u_{n,\tau,2}(x)| \leq c_n\}.
$$

For example, the set $B_{n,\{1\}}(c_n)$ is a set of points (x,τ) such that $|v_{n,\tau,1}(x)/\sigma_{n,\tau,1}(x)|$ is close to zero, and $v_{n,\tau,2}(x)/\sigma_{n,\tau,2}(x)$ is negative and away from zero. We call *contact sets* such sets as $B_{n,\{1\}}(c_n)$, $B_{n,\{2\}}(c_n)$, and $B_{n,\{1,2\}}(c_n)$.

Now, comparing (2.3) with (2.1) reveals that the limiting distribution of $\hat{\theta}$ under the null hypothesis will not depend on points outside the union of the contact sets. Thus it is natural to base the bootstrap critical values on the quantity on the right hand side of (2.3) instead of that on the last integral in (2.1). As we will explain shortly in the next subsection, this leads to a test that is uniformly valid and exhibits substantial improvement in power.

To construct bootstrap critical values, we introduce sample versions of the contact sets:

$$
\hat{B}_{\{1\}}(c_n) \equiv \{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |\hat{u}_{\tau,1}(x)| \leq c_n \text{ and } \hat{u}_{\tau,2}(x) < -c_n\},
$$
$$
\hat{B}_{\{2\}}(c_n) \equiv \{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |\hat{u}_{\tau,2}(x)| \leq c_n \text{ and } \hat{u}_{\tau,1}(x) < -c_n\} \text{ and}
$$
$$
\hat{B}_{\{1,2\}}(c_n) \equiv \{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |\hat{u}_{\tau,1}(x)| \leq c_n \text{ and } |\hat{u}_{\tau,2}(x)| \leq c_n\}.
$$

See Figure 1 for illustration of estimation of contact sets when $J = 2$.
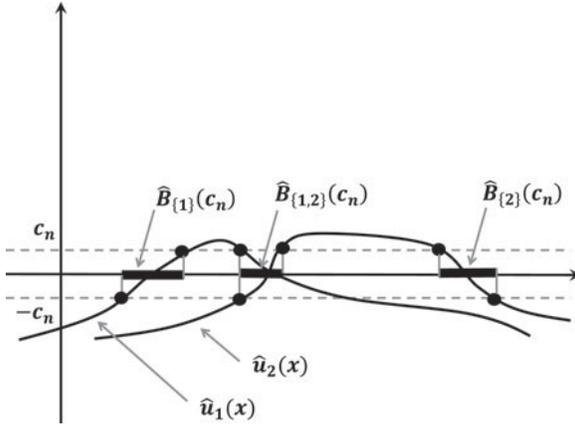
FIGURE 1. Contact set estimation.

Note: This figure illustrates estimated contact sets when $J = 2$. The thick line segments on the x-axis represent estimated contact sets.

Given the contact sets, we construct a bootstrap version of the right hand side of (2.3) as

$$\hat{\theta}^* \equiv \int_{\hat{B}_{\{1\}}(\hat{c}_n)} \max\left\{\hat{s}^*_{\tau,1}(x),0\right\}^p dQ(x,\tau)$$

$$+ \int_{\hat{B}_{\{2\}}(\hat{c}_n)} \max\left\{\hat{s}^*_{\tau,2}(x),0\right\}^p dQ(x,\tau)$$

$$+ \int_{\hat{B}_{\{1,2\}}(\hat{c}_n)} \max\left\{\hat{s}^*_{\tau,1}(x),\hat{s}^*_{\tau,2}(x),0\right\}^p dQ(x,\tau), \tag{2.4}$$

where $\hat{c}_n$ is a data dependent version of c_n. We will discuss a way to construct $\hat{c}_n$ shortly. We also define

$$\hat{a}^* \equiv \mathbf{E}^*\hat{\theta}^*,$$

where $\mathbf{E}^*$ denotes the expectation under the bootstrap distribution. Let c^*_α be the $(1-\alpha)$-th quantile from the bootstrap distribution of $\hat{\theta}^*$. In practice, both quantities $\hat{a}^*$ and c^*_α are approximated by the sample mean and sample $(1-\alpha)$-th quantile, respectively, from a large number of bootstrap repetitions. Then for a small constant $\eta \equiv 10^{-3}$, we take $c^*_{\alpha,\eta} \equiv \max\{c^*_\alpha, h^{d/2}\eta + \hat{a}^*\}$ as the critical value to form the following test:

Reject H_0 if and only if $\hat{\theta} > c^*_{\alpha,\eta}$. \tag{2.5}

Then it is shown later that the test has asymptotically correct size, i.e.,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} P\{\hat{\theta} > c^*_{\alpha,\eta}\} \leq \alpha, \tag{2.6}$$

where $\mathcal{P}_0$ is the collection of potential distributions that satisfy the null hypothesis. To implement the test, there are two important tuning parameters, namely the bandwidth h used for nonparametric estimation and the constant $\hat{c}_n$ for contact set estimation.[4] We discuss how to obtain the latter in the context of our Monte Carlo experiments in Section 5.1.

2.3. Discontinuity, Uniformity, and Power

Many tests of inequality restrictions exhibit discontinuity in its limiting distribution under the null hypothesis. When the inequality restrictions involve nonparametric functions, this discontinuity takes a complex form, as emphasized in Section 5 of Andrews and Shi (2013).

To see the discontinuity problem in our context, let $\{(Y_i, X_i)^\top\}_{i=1}^n$ be i.i.d. copies from an observable bivariate random vector, $(Y, X)^\top \in \mathbf{R} \times \mathbf{R}$, where X_i is a continuous random variable with density f. We consider a simple testing example:

$$H_0 : \mathbf{E}[Y|X = x] \leq 0 \text{ for all } x \in \mathcal{X} \text{ vs. } H_1 : \mathbf{E}[Y|X = x] > 0 \text{ for some } x \in \mathcal{X}.$$
(2.7)

Here, with the subscript τ suppressed, we set $J = 1$, $r_{n,1} = \sqrt{nh}$, $p = d = 1$, and define $[v]_+ \equiv \max\{v, 0\}$. Let

$$\hat{v}_1(x) = \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right) \text{ and } \hat{\sigma}_1^2(x) = \frac{1}{nh} \sum_{i=1}^n Y_i^2 K^2\left(\frac{X_i - x}{h}\right),$$
(2.8)

where K is a nonnegative, univariate kernel function with compact support and h is a bandwidth.

Assume that the density of X is strictly positive on $\mathcal{X}$. Then, in this example, $v_{n,1}(x) \equiv \mathbf{E}\hat{v}_1(x) \leq 0$ for almost every x in $\mathcal{X}$ whenever the null hypothesis is true. Define

$$Z_{n,1}(x) = \sqrt{nh}\left\{\frac{\hat{v}_1(x) - v_{n,1}(x)}{\hat{\sigma}_1(x)}\right\} \text{ and } B_{n,1}(0) = \left\{x \in \mathcal{X} : \left|\sqrt{nh}v_{n,1}(x)\right| = 0\right\}.$$

We analyze the asymptotic properties of $\hat{\theta}$ as follows. We first write

$$h^{-1/2}(\hat{\theta} - a_{n,1}) = h^{-1/2}\left\{\int_{B_{n,1}(0)} [Z_{n,1}(x)]_+\, dx - a_{n,1}\right\}$$
(2.9)

$$+ h^{-1/2}\int_{\mathcal{X}\backslash B_{n,1}(0)} \left[Z_{n,1}(x) + \frac{\sqrt{nh}v_{n,1}(x)}{\hat{\sigma}_1(x)}\right]_+\, dx,$$

where

$$a_{n,1} = \mathbf{E}\left[\int_{B_{n,1}(0)} [Z_{n,1}(x)]_+\, dx\right].$$

When $\liminf_{n\to\infty} Q(B_{n,1}(0)) > 0$ with $Q(B_{n,1}(0))$ denoting Lebesgue measure of $B_{n,1}(0)$, we can show that the leading term on the right hand side in (2.9) becomes asymptotically $N(0, \sigma_0^2)$ for some $\sigma_0^2 > 0$. On the other hand, the second term vanishes in probability as $n \to \infty$ under H_0 because for each $x \in \mathcal{X}\backslash B_{n,1}(0)$,

$$0 > \sqrt{nh} v_{n,1}(x) \to -\infty$$

as $n \to \infty$ under H_0. Thus we conclude that when $\liminf_{n\to\infty} Q(B_{n,1}(0)) > 0$ under H_0,

$$h^{-1/2}(\hat{\theta} - a_{n,1}) \approx h^{-1/2}\left\{ \int_{B_{n,1}(0)} [Z_{n,1}(x)]_+ \, dx - a_{n,1} \right\} \to_d N(0, \sigma_0^2). \quad \textbf{(2.10)}$$

This asymptotic theory is pointwise in P (with P fixed and letting $n \to \infty$), and may not be adequate for finite sample approximation. There are two sources of discontinuity. First, the pointwise asymptotic theory essentially regards the drift component $\sqrt{nh} v_{n,1}(x)$ as $-\infty$, whereas in finite samples, the component can be very negative, but not $-\infty$. Second, even if the nonparametric function $\sqrt{nh} v_{n,1}(x)$ changes continuously, the contact set $B_{n,1}(0)$ may change discontinuously in response.[5] While there is no discontinuity in the finite sample distribution of the test statistic, there may arise discontinuity in its pointwise asymptotic distribution. Furthermore, the complexity of the discontinuity makes it harder to trace its source, when we have $J > 2$. As a result, the asymptotic validity of the test that is established pointwise in P is not a good justification of the test. We need to establish the asymptotic validity that is *uniform* in P over a reasonable class of probabilities.

Recall that bootstrap critical values based on the least favorable configuration use a bootstrap quantity such as

$$\hat{\theta}^*_{\text{LFC}} \equiv \int_{\mathcal{X}} [\hat{s}^*(x)]_+ \, dx, \text{ where } \hat{s}^*(x) = \sqrt{nh}\left\{ \frac{\hat{v}_1^*(x) - \hat{v}_1(x)}{\hat{\sigma}_1^*(x)} \right\}, \quad \textbf{(2.11)}$$

which can yield tests that are asymptotically valid *uniformly* in P. However, using a critical value based on

$$\hat{\theta}_1^* \equiv \int_{\hat{B}_{\{1\}}(c_n)} [\hat{s}^*(x)]_+ \, dx$$

also yields an asymptotically valid test, and yet $\hat{\theta}^*_{\text{LFC}} > \hat{\theta}_1^*$ in general. Thus the bootstrap tests that use the contact set have better power properties than those that do not. The power improvement is substantial in many simulation designs and can be important in real-data applications.

Now, let us see how the choice of $c^*_{\alpha,\eta} \equiv \max\{c^*_\alpha, h^{1/2}\eta + \hat{a}^*\}$ (with $d = 1$ here) leads to bootstrap inference that is valid even when the test statistic becomes degenerate under the null hypothesis. The degeneracy arises when the inequality

restrictions hold with large slackness, so that the convergence in (2.10) holds with $\sigma_0^2 = 0$, and hence

$$h^{-1/2}(\hat{\theta} - a_{n,1}) = o_P(1).$$

For the bootstrap counterpart, note that

$$h^{-1/2}(c_{\alpha,\eta}^* - a_{n,1}) = h^{-1/2} \max\{c_\alpha^* - a_{n,1}, h^{1/2}\eta + \hat{a}^* - a_{n,1}\}$$

$$\geq \eta + h^{-1/2}(\hat{a}^* - a_{n,1}),$$

where it can be shown that $h^{-1/2}(\hat{a}^* - a_{n,1}) = o_P(1)$. Therefore, the bootstrap inference is designed to be asymptotically valid even when the test statistic becomes degenerate.

Note that for the sake of validity only, one may replace $h^{1/2}\eta$ by a fixed constant, say $\bar{\eta} > 0$. However, this choice would render the test asymptotically too conservative. The choice of $h^{1/2}\eta$ in this article makes the test asymptotically exact for a wide class of probabilities, while preserving the uniform validity in both the cases of degeneracy and nondegeneracy.[6] The precise class of probabilities under which the test becomes asymptotically exact is presented in Section 3.

There are two remarkable aspects of the local power behavior of our bootstrap test. First, the test exhibits two different kinds of convergence rates along different directions of Pitman local alternatives. Second, despite the fact that the test uses the approach of local smoothing by kernel as in Härdle and Mammen (1993), the faster of the two convergence rates achieves a parametric rate of $\sqrt{n}$. To see this more closely, let us return to the simple example in (2.7), and consider the following local alternatives:

$$\upsilon_n(x) = \upsilon_0(x) + \frac{\delta(x)}{b_n}, \tag{2.12}$$

where $\upsilon_0(x) \leq 0$ for all $x \in \mathcal{X}$ and $\delta(x) > 0$ for some $x \in \mathcal{X}$, and $b_n \to \infty$ as $n \to \infty$ such that $\upsilon_n(x) > 0$ for some $x \in \mathcal{X}$. The function $\delta(\cdot)$ represents a Pitman direction of the local alternatives. As we show later, there exist two types of convergence rates of our test, depending on the choice of $\delta(x)$. Let $B^0(0) \equiv \{x \in \mathcal{X} : \upsilon_0(x) = 0\}$ and $\sigma_1^2(x) \equiv \mathbf{E}[Y_i^2|X_i = x]f(x) \int K^2(u)du$. When $\delta(\cdot)$ is such that

$$\int_{B^0(0)} \frac{\delta(x)}{\sigma_1(x)}dx > 0,$$

the test achieves a parametric rate $b_n = \sqrt{n}$. On the other hand, when $\delta(\cdot)$ is such that

$$\int_{B^0(0)} \frac{\delta(x)}{\sigma_1(x)}dx = 0 \text{ and } \int_{B^0(0)} \frac{\delta^2(x)}{\sigma_1^2(x)}dx > 0,$$

the test achieves a slower rate $b_n = \sqrt{n}h^{1/4}$. See Section 4.2 for heuristics behind the results. In Section 4.3, the general form of local power functions is derived.

3. UNIFORM ASYMPTOTICS UNDER GENERAL CONDITIONS

In this section, we establish uniform asymptotic validity of our bootstrap test.[7] We also provide a class of distributions for which the asymptotic size is exact. We first define the set of distributions we consider.

DEFINITION 1. *Let $\mathcal{P}$ denote the collection of the potential joint distributions of the observed random vectors that satisfy Assumptions A1–A6, and B1–B4 given below. Let $\mathcal{P}_0 \subset \mathcal{P}$ be the sub-collection of potential distributions that satisfy the null hypothesis.*

Let $\|\cdot\|$ denote the Euclidean norm throughout the article. For any given sequence of subcollections $\mathcal{P}_n \subset \mathcal{P}$, any sequence of real numbers $b_n > 0$, and any sequence of random vectors Z_n, we say that $Z_n/b_n \to_P 0$, $\mathcal{P}_n$-uniformly, or $Z_n = o_P(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\{\|Z_n\| > ab_n\} = 0.$$

Similarly, we say that $Z_n = O_P(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\{\|Z_n\| > Mb_n\} < a.$$

We also define their bootstrap counterparts. Let P^* denote the probability under the bootstrap distribution. For any given sequence of subcollections $\mathcal{P}_n \subset \mathcal{P}$, any sequence of real numbers $b_n > 0$, and any sequence of random vectors Z_n^*, we say that $Z_n^*/b_n \to_{P^*} 0$, $\mathcal{P}_n$-uniformly, or $Z_n^* = o_{P^*}(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\{P^*\{\|Z_n^*\| > ab_n\} > a\} = 0.$$

Similarly, we say that $Z_n^* = O_{P^*}(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\{P^*\{\|Z_n^*\| > Mb_n\} > a\} < a.$$

In particular, when we say $Z_n = o_P(b_n)$ or $O_P(b_n)$, $\mathcal{P}$-uniformly, it means that the convergence holds uniformly over $P \in \mathcal{P}$, and when we say $Z_n = o_P(b_n)$ or $O_P(b_n)$, $\mathcal{P}_0$-uniformly, it means that the convergence holds uniformly over all the probabilities in $\mathcal{P}$ that satisfy the null hypothesis.

3.1. Test Statistics and Critical Values in General Form

First, let us extend the test statistics and the bootstrap procedure to the general case of $J \geq 1$. Let $\Lambda_p : \mathbf{R}^J \to [0, \infty)$ be a nonnegative, increasing function indexed by

p such that $1 \leq p < \infty$. While the theory of this article can be extended to various general forms of map Λ_p, we focus on the following type:

$$\Lambda_p(v_1, \ldots, v_J) = (\max\{[v_1]_+, \ldots, [v_J]_+\})^p \text{ or } \Lambda_p(v_1, \ldots, v_J) = \sum_{j=1}^{J} [v_j]_+^p, \quad \text{(3.1)}$$

where for $a \in \mathbf{R}$, $[a]_+ = \max\{a, 0\}$. The test statistic is defined as

$$\hat{\theta} = \int_{\mathcal{X} \times \mathcal{T}} \Lambda_p\left(\hat{u}_{\tau,1}(x), \ldots, \hat{u}_{\tau,J}(x)\right) dQ(x, \tau).$$

To motivate our bootstrap procedure, it is convenient to begin with the following lemma. Let us introduce some notation. Define $\mathcal{N}_J \equiv 2^{\mathbb{N}_J} \backslash \{\varnothing\}$, i.e., the collection of all the nonempty subsets of $\mathbb{N}_J \equiv \{1, 2, \ldots, J\}$. For any $A \in \mathcal{N}_J$ and $\mathbf{v} = (v_1, \ldots, v_J)^\top \in \mathbf{R}^J$, we define $\mathbf{v}_A$ to be $\mathbf{v}$ except that for each $j \in \mathbb{N}_J \backslash A$, the j-th entry of $\mathbf{v}_A$ is zero, and let

$$\Lambda_{A,p}(\mathbf{v}) \equiv \Lambda_p(\mathbf{v}_A). \quad \text{(3.2)}$$

That is, $\Lambda_{A,p}(\mathbf{v})$ is a "censoring" of $\Lambda_p(\mathbf{v})$ outside the index set A. Now, we define a general version of contact sets: for $A \in \mathcal{N}_J$ and for $c_{n,1}, c_{n,2} > 0$,

$$B_{n,A}(c_{n,1}, c_{n,2}) \equiv \left\{ (x, \tau) \in \mathcal{X} \times \mathcal{T} : \begin{array}{l} |r_{n,j} v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)| \leq c_{n,1}, \text{ for all } j \in A \\ r_{n,j} v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x) < -c_{n,2}, \text{ for all } j \in \mathbb{N}_J/A \end{array} \right\},$$
$$\text{(3.3)}$$

where $\sigma_{n,\tau,j}(x)$ is a "population" version of $\hat{\sigma}_{\tau,j}(x)$ (see e.g., Assumption A5 below). When $c_{n,1} = c_{n,2} = c_n$ for some $c_n > 0$, we write $B_{n,A}(c_n) = B_{n,A}(c_{n,1}, c_{n,2})$.

LEMMA 1. *Suppose that Assumptions A1–A3 and A4(i) in Section 3.2 hold. Suppose further that $c_{n,1} > 0$ and $c_{n,2} > 0$ are sequences such that*

$$\sqrt{\log n}\{c_{n,1}^{-1} + c_{n,2}^{-1}\} \to 0,$$

as $n \to \infty$. Then as $n \to \infty$,

$$\inf_{P \in \mathcal{P}_0} P\left\{ \hat{\theta} = \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,1}, c_{n,2})} \Lambda_{A,p}(\hat{u}_{\tau,1}(x), \ldots, \hat{u}_{\tau,J}(x)) dQ(x, \tau) \right\} \to 1,$$

where $\mathcal{P}_0$ is the set of potential distributions of the observed random vector under the null hypothesis.

The lemma above shows that the test statistic $\hat{\theta}$ is uniformly approximated by the integral with domain restricted to the contact sets $B_{n,A}(c_{n,1}, c_{n,2})$ in large samples. Note that the result of Lemma 1 implies that the approximation error

between $\hat{\theta}$ and the expression on the right-hand side is $o_P(\varepsilon_n)$ for any $\varepsilon_n \to 0$, thereby suggesting that one may consider a bootstrap procedure that mimics the representation of $\hat{\theta}$ in Lemma 1.

We begin by introducing a sample version of the contact sets. For $A \in \mathcal{N}_J$,

$$\hat{B}_A(\hat{c}_n) \equiv \left\{ (x,\tau) \in \mathcal{X} \times \mathcal{T} : \begin{array}{ll} |r_{n,j}\hat{v}_{\tau,j}(x)/\hat{\sigma}_{\tau,j}(x)| \leq \hat{c}_n, & \text{for all } j \in A \\ r_{n,j}\hat{v}_{\tau,j}(x)/\hat{\sigma}_{\tau,j}(x) < -\hat{c}_n, & \text{for all } j \in \mathbb{N}_J \backslash A \end{array} \right\}.$$

The explicit condition for $\hat{c}_n$ is found in Assumption A4 below. Given the bootstrap counterparts, $\{[\hat{v}^*_{\tau,j}(x), \hat{\sigma}^*_{\tau,j}(x)] : j \in \mathbb{N}_J\}$, of $\{[\hat{v}_{\tau,j}(x), \hat{\sigma}_{\tau,j}(x)] : j \in \mathbb{N}_J\}$, we define our bootstrap test statistic as follows:

$$\hat{\theta}^* \equiv \sum_{A \in \mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \Lambda_{A,p}(\hat{s}^*_{\tau,1}(x), \dots, \hat{s}^*_{\tau,J}(x)) dQ(x,\tau),$$

where for $j \in \mathbb{N}_J$, $\hat{s}^*_{\tau,j}(x) \equiv r_{n,j}(\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x))/\hat{\sigma}^*_{\tau,j}(x)$. We also define

$$\hat{a}^* \equiv \sum_{A \in \mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \mathbf{E}^* \Lambda_{A,p}(\hat{s}^*_{\tau,1}(x), \dots, \hat{s}^*_{\tau,J}(x)) dQ(x,\tau).$$

Let c^*_α be the $(1-\alpha)$-th quantile from the bootstrap distribution of $\hat{\theta}^*$ and take

$$c^*_{\alpha,\eta} = \max\{c^*_\alpha, h^{d/2}\eta + \hat{a}^*\}$$

as our critical value, where $\eta \equiv 10^{-3}$ is a small fixed number.

One of the main technical contributions of this article is to present precise conditions under which this proposal of bootstrap test works. We present and discuss them in subsequent sections.

To see the intuition for the bootstrap validity, first note that the uniform convergence of $r_{n,j}\{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)\}/\hat{\sigma}_{\tau,j}(x)$ over (x,τ) implies that

$$B_{n,A}(c_{n,L}, c_{n,U}) \subset \hat{B}_A(\hat{c}_n) \subset B_{n,A}(c_{n,U}, c_{n,L}) \tag{3.4}$$

with probability approaching one, whenever $P\{c_{n,L} \leq \hat{c}_n \leq c_{n,U}\} \to 1$. Therefore, if $\sqrt{\log n}/c_{n,L} \to 0$, then, (letting $\hat{s}_{\tau,j} \equiv r_{n,j}(\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x))/\hat{\sigma}_{\tau,j}(x)$), we have

$$\hat{\theta} \leq \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,L}, c_{n,U})} \Lambda_{A,p}(\hat{s}_{\tau,1}(x), \dots, \hat{s}_{\tau,J}(x)) dQ(x,\tau), \tag{3.5}$$

with probability approaching one, by Lemma 1 and the null hypothesis. When the last sum has a nondegenerate limit, we can approximate its distribution by the bootstrap distribution

$$\sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,L}, c_{n,U})} \Lambda_{A,p}(\hat{s}^*_{\tau,1}(x), \dots, \hat{s}^*_{\tau,J}(x)) dQ(x,\tau)$$

$$\leq \sum_{A \in \mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \Lambda_{A,p}(\hat{s}^*_{\tau,1}(x), \dots, \hat{s}^*_{\tau,J}(x)) dQ(x,\tau) \equiv \hat{\theta}^*,$$

where the inequality follows from (3.4).[8] Thus the critical value is read from the bootstrap distribution of $\hat{\theta}^*$. On the other hand, if the last sum in (3.5) has limiting distribution degenerate at zero, we simply take a small positive number η to control the size of the test. This results in our choice of $c_{\alpha,\eta}^* = \max\{c_{\alpha}^*, h^{d/2}\eta + \hat{a}^*\}$.

3.2. Assumptions

In this section, we provide assumptions needed to develop general results. We assume that $\mathcal{S} \equiv \mathcal{X} \times \mathcal{T}$ is a compact subset of a Euclidean space. We begin with the following assumption.

Assumption A1 (Asymptotic linear representation). For each $j \in \mathbb{N}_J \equiv \{1, \ldots, J\}$, there exists a nonstochastic function $v_{n,\tau,j}(\cdot) : \mathbf{R}^d \to \mathbf{R}$ such that (a) $v_{n,\tau,j}(x) \leq 0$ for all $(x, \tau) \in \mathcal{S}$ under the null hypothesis, and (b) as $n \to \infty$,

$$\sup_{(x,\tau)\in\mathcal{S}} \left| r_{n,j}\left\{ \frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right\} - \sqrt{nh^d}\{\hat{g}_{\tau,j}(x) - \mathbf{E}\hat{g}_{\tau,j}(x)\} \right|$$
$$= o_P(\sqrt{h^d}), \; \mathcal{P}\text{-uniformly}, \tag{3.6}$$

where, with $\{(Y_i^\top, X_i^\top)\}_{i=1}^n$ being a random sample such that $Y_i = (Y_{i1}^\top, \ldots, Y_{iJ}^\top)^\top \in \mathbf{R}^{J\bar{L}}$, $Y_{ij} \in \mathbf{R}^{\bar{L}}$, $X_i \in \mathbf{R}^d$, and the distribution of X_i is absolutely continuous with respect to Lebesgue measure,[9] we define

$$\hat{g}_{\tau,j}(x) \equiv \frac{1}{nh^d} \sum_{i=1}^n \beta_{n,x,\tau,j}\left(Y_{ij}, \frac{X_i - x}{h}\right),$$

and $\beta_{n,x,\tau,j} : \mathbf{R}^{\bar{L}} \times \mathbf{R}^d \to \mathbf{R}$ is a function which may depend on $n \geq 1$.

Assumption A1 requires that there exist a nonparametric function $v_{n,\tau,j}(x)$ around which the asymptotic linear representation holds uniformly in $P \in \mathcal{P}$, and $v_{n,\tau,j}(x) \leq 0$ under the null hypothesis. The required rate of convergence in (3.6) is $o_P(h^{d/2})$ instead of $o_P(1)$. We need this stronger convergence rate primarily because $\hat{\theta} - a_n$ is $O_P(h^{d/2})$ for some nonstochastic sequence a_n.[10]

When $\hat{v}_{\tau,j}(x)$ is a sample mean of i.i.d. random quantities involving non-negative kernels and $\hat{\sigma}_{n,\tau}(x) = 1$, we may take $v_{n,\tau,j}(x) = \mathbf{E}\hat{v}_{\tau,j}(x)$, and then $o_P(\sqrt{h^d})$ is in fact precisely equal to 0. If the original nonparametric function $v_{\tau,j}(\cdot)$ satisfies some smoothness conditions, we may take $v_{n,\tau,j}(x) = v_{\tau,j}(x)$, and handle the bias part $\mathbf{E}\hat{v}_{\tau,j}(x) - v_{\tau,j}(x)$ using the standard arguments to deduce the error rate $o_P(\sqrt{h^d})$. Assumption A1 admits both set-ups. For instance, consider the simple example in Section 2.3. The asymptotic linear representation in Assumption A1 can be shown to hold with

$$\beta_{n,x,1}(Y_i, (X_i - x)/h) = Y_i K((X_i - x)/h)/\sigma_{n,1}(x),$$

where $\sigma_{n,1}^2(x) = \mathbf{E}[Y_i^2 K^2((X_i - x)/h)]/h$, if $\hat{\sigma}_{n,1}(x)$ is chosen as in (2.8).

The following assumption for $\beta_{n,x,\tau,j}$ essentially defines the scope of this paper's framework.

Assumption A2 (Kernel-type condition). For some compact $\mathcal{K}_0 \subset \mathbf{R}^d$ that does not depend on $P \in \mathcal{P}$ or n, it is satisfied that $\beta_{n,x,\tau,j}(y, u) = 0$ for all $u \in \mathbf{R}^d \setminus \mathcal{K}_0$ and all $(x, \tau, y) \in \mathcal{X} \times \mathcal{T} \times \mathcal{Y}_j$ and all $j \in \mathbb{N}_J$, where $\mathcal{Y}_j$ denotes the support of Y_{ij}.

Assumption A2 can be immediately verified when the asymptotic linear representation in (3.6) is established. This condition is satisfied in particular when the asymptotic linear representation involves a multivariate kernel function with bounded support in a multiplicative form. In such a case, the set $\mathcal{K}_0$ depends only on the choice of the kernel function, not on any model primitives.

Assumption A3 (Uniform convergence rate for nonparametric estimators). For all $j \in \mathbb{N}_J$,

$$\sup_{(x,\tau) \in \mathcal{S}} r_{n,j} \left| \frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right| = O_P\left(\sqrt{\log n}\right), \ \mathcal{P}\text{-uniformly.}$$

Assumption A3 (in combination with A5 below) requires that $\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)$ have the uniform convergence rate of $O_P(r_{n,j}^{-1}\sqrt{\log n})$ uniformly over $P \in \mathcal{P}$. Lemma 2 in Section 3.4 provides some sufficient conditions for this convergence.

We now introduce conditions for the bandwidth h and the tuning parameter c_n for the contact sets.

Assumption A4 (Rate conditions for tuning parameters).   (i) As $n \to \infty$, $h \to 0$, $\sqrt{\log n}/r_n \to 0$, and $n^{-1/2}h^{-d-v_1} \to 0$ for some arbitrarily small $v_1 > 0$, where $r_n \equiv \min_{j \in \mathbb{N}_J} r_{n,j}$.
(ii) For each $n \geq 1$, there exist nonstochastic sequences $c_{n,L} > 0$ and $c_{n,U} > 0$ such that $c_{n,L} \leq c_{n,U}$, and

$$\inf_{P \in \mathcal{P}} P\left\{ c_{n,L} \leq \hat{c}_n \leq c_{n,U} \right\} \to 1, \text{ and } \sqrt{\log n}/c_{n,L} + c_{n,U}/r_n \to 0,$$

as $n \to \infty$.

The requirement that $\sqrt{\log n}/r_n \to 0$ is satisfied easily for most cases where r_n increases at a polynomial order in n. Assumption A4(ii) requires that $\hat{c}_n$ increase faster than $\sqrt{\log n}$ but slower than r_n with probability approaching one.

Assumption A5 (Regularity conditions for $\hat{\sigma}_{\tau,j}(x)$). For each $(\tau, j) \in \mathcal{T} \times \mathbb{N}_J$, there exists $\sigma_{n,\tau,j}(\cdot) : \mathcal{X} \to (0, \infty)$ such that $\liminf_{n \to \infty} \inf_{(x,\tau) \in \mathcal{S}} \inf_{P \in \mathcal{P}} \sigma_{n,\tau,j}(x) > 0$, and

$$\sup_{(x,\tau) \in \mathcal{S}} \left| \hat{\sigma}_{\tau,j}(x) - \sigma_{n,\tau,j}(x) \right| = o_P(1), \ \mathcal{P}\text{-uniformly.}$$

Assumption A5 requires that the scale normalization $\hat{\sigma}_{\tau,j}(x)$ should be asymptotically well defined. The condition precludes the case where estimator $\hat{\sigma}_{\tau,j}(x)$

converges to a map that becomes zero at some point (x, τ) in $\mathcal{S}$. Assumption A5 is usually satisfied by an appropriate choice of $\hat{\sigma}_{\tau, j}(x)$. When one chooses $\hat{\sigma}_{\tau, j}(x) = 1$, which is permitted in our framework, Assumption A5 is immediately satisfied with $\sigma_{n, \tau, j}(x) = 1$. Again, if we go back to the simple example considered in Section 2.3, it is straightforward to see that under regularity conditions, with the subscript τ suppressed, $\hat{\sigma}_1^2(x) = \sigma_{n,1}^2(x) + o_P(1)$ and $\sigma_{n,1}^2(x) = \sigma_1^2(x) + o(1)$, where $\sigma_1^2(x) \equiv \mathbf{E}(Y^2 | X = x) f(x) \int K^2(u) du$, as $n \to \infty$. The convergence can be strengthened to a uniform convergence when $\sigma_1^2(x)$ is bounded away from zero uniformly over $x \in \mathcal{X}$ and $P \in \mathcal{P}$, so that Assumption A5 holds.

We introduce assumptions about the moment conditions for $\beta_{n,x,\tau,j}(\cdot, \cdot)$ and other regularity conditions. For $\tau \in \mathcal{T}$ and $\varepsilon_1 > 0$, let $\mathcal{S}_\tau(\varepsilon_1) \equiv \{x + a : x \in \mathcal{S}_\tau, a \in [-\varepsilon_1, \varepsilon_1]^d\}$, where $\mathcal{S}_\tau \equiv \{x \in \mathcal{X} : (x, \tau) \in \mathcal{S}\}$ for each $\tau \in \mathcal{T}$. Let $\mathcal{U} \equiv \mathcal{K}_0 + \mathcal{K}_0$ such that $\mathcal{U}$ contains $\{0\}$ in its interior and $\mathcal{K}_0$ is the same as Assumption A2. Here, $+$ denotes the Minkowski sum of sets.

Assumption A6.   (i) There exist $M \geq 2(p+2)$, $C > 0$, and $\varepsilon_1 > 0$ such that

$$\mathbf{E}[|\beta_{n,x,\tau,j}(Y_{ij}, u)|^M | X_i = x] f(x) \leq C,$$

for all $(x, u) \in \mathcal{S}_\tau(\varepsilon_1) \times \mathcal{U}$, $\tau \in \mathcal{T}$, $j \in \mathbb{N}_J$, $n \geq 1$, and $P \in \mathcal{P}$, where $f(\cdot)$ is the density of X_i.[11]

(ii) For each $a \in (0, 1/2)$, there exists a compact set $\mathcal{C}_a \subset \mathbf{R}^d$ such that

$$0 < \inf_{P \in \mathcal{P}} P\{X_i \in \mathbf{R}^d \backslash \mathcal{C}_a\} \leq \sup_{P \in \mathcal{P}} P\{X_i \in \mathbf{R}^d \backslash \mathcal{C}_a\} < a.$$

Assumption A6(i) requires that conditional moments of $\beta_{n,x,\tau,j}(Y_{ij}, z)$ be bounded. Assumption A6(ii) is a technical condition for the distribution of X_i. The third inequality in Assumption A6(ii) is satisfied if the distribution of X_i is uniformly tight in $\mathcal{P}$, and follows, for example, if $\sup_{P \in \mathcal{P}} \mathbf{E} \| X_i \| < \infty$. The first inequality in Assumption A6(ii) requires that there be a common compact set outside which the distribution of X_i still has positive probability mass uniformly over $P \in \mathcal{P}$. The main thrust of Assumption A6(ii) lies in the requirement that such a compact set be independent of $P \in \mathcal{P}$. While it is necessary to make this technical condition explicit as stated here, the condition itself appears very weak.

This article asymptotic analysis adopts the approach of Poissonization (see, e.g., Horváth (1991) and Giné, Mason, and Zaitsev (2003)). However, existing methods of Poissonization are not readily applicable to our testing problem, mainly due to the possibility of local or global redundancy among the nonparametric functions. In particular, the conditional covariance matrix of $\beta_{n,x,\tau,j}(Y_{ij}, u)$'s across different (x, τ, j)'s given X_i can be singular in the limit. Since the empirical researcher rarely knows *a priori* the local relations among nonparametric functions, it is important that the validity of the test is not sensitive to the local relations among them, i.e., the validity should be uniform in P.

This article deals with this challenge in three steps. First, we introduce a Poissonized version of the test statistic and apply a certain form of regularization

to facilitate the derivation of its limiting distribution uniformly in $P \in \mathcal{P}$, i.e., regardless of singularity or degeneracy in the original test statistic. Second, we use a Berry-Esseen-type bound to compute the finite sample influence of the regularization bias and let the regularization parameter go to zero carefully, so that the bias disappears in the limit. Third, we translate thus computed limiting distribution into that of the original test statistic, using so-called de-Poissonization lemma. This is how the uniformity issue in this complex situation is covered through the Poissonization method combined with the method of regularization.

3.3. Asymptotic Validity of Bootstrap Procedure

Recall that $\mathbf{E}^*$ and P^* denote the expectation and the probability under the bootstrap distribution. We make the following assumptions for $\hat{v}^*_{\tau,j}(x)$.

Assumption B1 (Bootstrap asymptotic linear representation). For each $j \in \mathbb{N}_J$,

$$\sup_{(x,\tau)\in\mathcal{S}} \left| r_{n,j} \left\{ \frac{\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x)}{\hat{\sigma}^*_{\tau,j}(x)} \right\} - \sqrt{nh^d} \{ \hat{g}^*_{\tau,j}(x) - \mathbf{E}^* \hat{g}^*_{\tau,j}(x) \} \right|$$
$$= o_{P^*}(\sqrt{h^d}), \ \mathcal{P}\text{-uniformly},$$

where

$$\hat{g}^*_{\tau,j}(x) \equiv \frac{1}{nh^d} \sum_{i=1}^{n} \beta_{n,x,\tau,j} \left(Y^*_{ij}, \frac{X^*_i - x}{h} \right),$$

and $\beta_{n,x,\tau,j}$ is a real valued function introduced in Assumption A1.

Assumption B2. For all $j \in \mathbb{N}_J$,

$$\sup_{(x,\tau)\in\mathcal{S}} r_{n,j} \left| \frac{\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x)}{\hat{\sigma}^*_{\tau,j}(x)} \right| = O_{P^*}(\sqrt{\log n}), \ \mathcal{P}\text{-uniformly}.$$

Assumption B3. For all $j \in \mathbb{N}_J$,

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \hat{\sigma}^*_{\tau,j}(x) - \hat{\sigma}_{\tau,j}(x) \right| = o_{P^*}(1), \ \mathcal{P}\text{-uniformly}.$$

Assumption B1 is the asymptotic linear representation of the bootstrap estimator $\hat{v}^*_{\tau,j}(x)$. The proof of the asymptotic linear representation can be typically proceeded in a similar way that one obtains the original asymptotic linear representation in Assumption A1. Assumptions B2 and B3 are the bootstrap versions of Assumptions A3 and A5. Proving Assumption B3 is very similar to the way we prove Assumption A5. One can use similar arguments in the proof of Lemma 2(ii) below. (See the supplemental note for its proof.) Sufficient conditions for Assumption B2 are provided in Lemma 2(ii) below.

Assumption B4 (Bandwidth condition). $n^{-1/2} h^{-\left(\frac{3M-4}{2M-4}\right)d - v_2} \to 0$ as $n \to \infty$, for some small $v_2 > 0$ and for $M > 0$ that appears in Assumption A6(i).

When $\beta_{n,x,\tau,j}(Y_{ij}, u)$ is bounded uniformly over (n, x, τ, j), the bandwidth condition in Assumption B4 can be reduced to $n^{-1/2}h^{-3d/2-v_2} \to 0$. If Assumption A6(i) holds with $M = 6$ and $p = 1$, the bandwidth condition in Assumption B4 is reduced to $n^{-1/2}h^{-7d/4-v_2} \to 0$.

Note that Assumption B4 is stronger than the bandwidth condition in Assumption A4(i). The main reason is that we need to prove that for some $a_\infty > 0$, we have $a_n = a_\infty + o(h^{d/2})$ and $a_n^* = a_\infty + o_P(h^{d/2})$, $\mathcal{P}$-uniformly, where a_n is an appropriate location normalizer of the test statistic, and a_n^* is a bootstrap counterpart of a_n. To show these, we utilize a Berry-Esseen-type bound for a nonlinear transform of independent sum of random variables. Since the approximation error depends on the moment bounds for the sum, the bandwidth condition in Assumption B4 takes a form that involves $M > 0$ in Assumption A6.

We now present the result of the uniform validity of our bootstrap test.

THEOREM 1. *Suppose that Assumptions A1–AA6 and B1–B4 hold. Then*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} P\{\hat{\theta} > c_{\alpha,\eta}^*\} \leq \alpha.$$

One might ask whether the bootstrap test $1\{\hat{\theta} > c_{\alpha,\eta}^*\}$ can be asymptotically exact, i.e., whether the inequality in Theorem 1 can hold as an equality. As we show below, the answer is affirmative. The remaining issue is a precise formulation of a subset of $\mathcal{P}_0$ such that the rejection probability of the bootstrap test achieves the level α asymptotically, uniformly over the subset.

To see when the test will have asymptotically exact size, we apply Lemma 1 to find that with probability approaching one,

$$\hat{\theta} = \sum_{A\in\mathcal{N}_J} \int_{B_{n,A}(c_{n,U}, c_{n,L})} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x; \hat{\sigma})\right) dQ(x, \tau),$$

where $\hat{\mathbf{s}}_\tau(x) \equiv \left[r_{n,j}\{\hat{v}_{n,\tau,j}(x) - v_{n,\tau,j}(x)\}/\hat{\sigma}_{\tau,j}(x)\right]_{j=1}^J$, and $\mathbf{u}_{n,\tau}(x; \hat{\sigma}) \equiv \left[r_{n,j}v_{n,\tau,j}(x)/\hat{\sigma}_{\tau,j}(x)\right]_{j=1}^J$, and $c_{n,U} > 0$ and $c_{n,L} > 0$ are nonstochastic sequences that satisfy Assumption A4(ii). We fix a positive sequence $q_n \to 0$, and write the right hand side as

$$\sum_{A\in\mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x; \hat{\sigma})\right) dQ(x, \tau)$$

$$+ \sum_{A\in\mathcal{N}_J} \int_{B_{n,A}(c_{n,U}, c_{n,L})\setminus B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x; \hat{\sigma})\right) dQ(x, \tau). \quad \textbf{(3.7)}$$

Under the null hypothesis, we have $v_{n,\tau,j}(x) \leq 0$, and hence the last sum is bounded by

$$\sum_{A\in\mathcal{N}_J} \int_{B_{n,A}(c_{n,U}, c_{n,L})\setminus B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right) dQ(x, \tau),$$

with probability approaching one. Using the uniform convergence rate in Assumption A3, we find that as long as

$$Q(B_{n,A}(c_{n,U}, c_{n,L}) \backslash B_{n,A}(q_n)) \to 0,$$

fast enough, the second term in (3.7) vanishes in probability. As for the first integral, since for all $x \in B_{n,A}(q_n)$, we have $|r_{n,j} v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)| \leq q_n$ for all $j \in A$, we use the Lipschitz continuity of the map $\Lambda_{A,p}$ on a compact set, to approximate the leading sum in (3.7) by

$$\bar{\theta}_{1,n}(q_n) \equiv \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right) dQ(x, \tau).$$

Thus we let

$$\tilde{\mathcal{P}}_n(\lambda_n, q_n) \equiv \left\{ P \in \mathcal{P} : Q\left(\bigcup_{A \in \mathcal{N}_J} B_{n,A}(c_{n,U}, c_{n,L}) \backslash B_{n,A}(q_n) \right) \leq \lambda_n \right\}, \tag{3.8}$$

and find that

$$\hat{\theta} = \bar{\theta}_{1,n}(q_n) + o_P(h^{d/2}), \ \tilde{\mathcal{P}}_n(\lambda_n, q_n) \cap \mathcal{P}_0\text{-uniformly},$$

as long as λ_n and q_n converge to zero fast enough. We will specify the conditions in Theorem 2 below.

Let us deal with $\bar{\theta}_{1,n}(q_n)$. First, it can be shown that there are sequences of nonstochastic numbers $a_n(q_n) \in \mathbf{R}$ and $\sigma_n(q_n) > 0$ that depend on q_n such that

$$h^{-d/2}\{\bar{\theta}_{1,n}(q_n) - a_n(q_n)\}/\sigma_n(q_n) \xrightarrow{d} N(0,1), \tag{3.9}$$

if $\liminf_{n \to \infty} \sigma_n(q_n) > 0$. We provide the precise formulae for $\sigma_n(q_n)$ and $a_n(q_n)$ in Section 4.3. Since the distribution of $h^{-d/2}\{\bar{\theta}_{1,n}(q_n) - a_n(q_n)\}/\sigma_n(q_n)$ is approximated by the bootstrap distribution of $h^{-d/2}\{\hat{\theta}^* - a_n(q_n)\}/\sigma_n(q_n)$ in large samples, we find that

$$\frac{h^{-d/2}\{c_\alpha^* - a_n(q_n)\}}{\sigma_n(q_n)} = \Phi^{-1}(1-\alpha) + o_P(1).$$

Hence the bootstrap critical value c_α^* will dominate $h^{-d/2}\eta + \hat{a}^* > 0$, if for all $n \geq 1$,

$$\Phi^{-1}(1-\alpha) \geq \frac{h^{-d/2}\{h^{d/2}\eta + \hat{a}^* - a_n(q_n)\}}{\sigma_n(q_n)}$$

$$= \frac{\eta + h^{-d/2}\{\hat{a}^* - a_n(q_n)\}}{\sigma_n(q_n)}.$$

We can show that $\hat{a}^* - a_n(q_n) = o_P(h^{d/2})$, which follows if λ_n in (3.8) vanishes to zero sufficiently fast. Hence if

$$\sigma_n(q_n) \geq \eta / \Phi^{-1}(1-\alpha),$$

we have c_α^* becomes approximately equal to our bootstrap critical value $c_{\alpha,\eta}^*$. This leads to the following formulation of probabilities.

DEFINITION 2. *Define*

$$\mathcal{P}_n(\lambda_n, q_n) \equiv \{P \in \tilde{\mathcal{P}}_n(\lambda_n, q_n) : \sigma_n(q_n) \geq \eta/\Phi^{-1}(1-\alpha)\},$$

where $\tilde{\mathcal{P}}_n(\lambda_n, q_n)$ *is as defined in* (3.8).

The following theorem establishes the asymptotic exactness of the size of the bootstrap test over $P \in \mathcal{P}_n(\lambda_n, q_n) \cap \mathcal{P}_0$.

THEOREM 2. *Suppose that Assumptions A1–A6 and B1–B4 hold. Let* $\lambda_n \to 0$ *and* $q_n \to 0$ *be positive sequences such that*

$$h^{-d/2}(\log n)^{p/2}\lambda_n \to 0 \text{ and} \tag{3.10}$$

$$h^{-d/2}q_n\{(\log n)^{(p-1)/2} + q_n^{p-1}\} \to 0.$$

Then

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n(\lambda_n,q_n)\cap\mathcal{P}_0} \left|P\{\hat{\theta} > c_{\alpha,\eta}^*\} - \alpha\right| = 0.$$

Theorem 2 shows that the rejection probability of our bootstrap test achieves exactly the level α uniformly over the set of probabilities in $\mathcal{P}_n(\lambda_n, q_n) \cap \mathcal{P}_0$. If $v_{n,\tau,j}(x) \equiv 0$ for each (x, τ) and for each j (the least favorable case, say P_{LFC}), then it is obvious that the distribution P_{LFC} belongs to $\mathcal{P}_n(\lambda_n, q_n)$ for any positive sequences $\lambda_n \to 0$ and $q_n \to 0$. This would be the only case of asymptotically exact coverage if bootstrap critical values were obtained as in (2.11), without contact set estimation. By estimating the contact sets and obtaining a critical value based on them, Theorem 2 establishes the asymptotically uniform exactness of the bootstrap test for distributions such that they may not satisfy $v_{n,\tau,j}(x) \equiv 0$ everywhere.

3.4. Sufficient Conditions for Uniform Convergences in Assumptions A3 and B2

This subsection gives sufficient conditions that yield Assumptions A3 and B2. The result is formalized in the following lemma.

LEMMA 2.   (i) *Suppose that Assumptions A1–A2 hold and that for each* $j \in \mathbb{N}_J$, *there exist finite constants* $C, \gamma_j > 0$, *and a positive sequence* $\delta_{n,j} > 0$ *such that for all* $n \geq 1$, *and all* $(x_1, \tau_1) \in \mathcal{S}$,

$$\mathbf{E}\left[\sup_{(x_2,\tau_2)\in\mathcal{S}:\|x_1-x_2\|+\|\tau_1-\tau_2\|\leq\lambda} \left(b_{n,ij}(x_1,\tau_1) - b_{n,ij}(x_2,\tau_2)\right)^2\right] \leq C\delta_{n,j}^2\lambda^{\gamma_j},$$

$$\text{for all } \lambda > 0, \quad (3.11)$$

where $b_{n,ij}(x_1,\tau_1) \equiv \beta_{n,x_1,\tau_1,j}\left(Y_{ij},(X_i-x_1)/h\right)$ *and* $\limsup_{n\to\infty}$ $\mathbf{E}[\sup_{(x,\tau)\in\mathcal{S}} b_{n,ij}^4(x,\tau)] \leq C$ *and* $\delta_{n,j} = n^{s_{1,j}}$ *and* $h = n^{s_2}$ *for some* $s_{1,j}, s_2 \in \mathbf{R}$. *Furthermore, assume that*

$$n^{-1/2}h^{-d-v} \to 0,$$

for some small $v > 0$. *Then, Assumption A3 holds.*

(ii) *Suppose further that Assumptions B1 and B3 hold. Then, Assumption B2 holds.*

The condition (3.11) is the local L_2-continuity condition for $\beta_{n,x,\tau,j}(Y_{ij},(X_i-x)/h)$ in (x,τ). The condition corresponds to what Andrews (1994) called "Type IV class". The condition is satisfied by numerous maps that are continuous or discontinuous, as long as regularity conditions for the random vector (Y_i, X_i) are satisfied.[12] Typically, $\delta_{n,j}$ diverges to infinity at a polynomial rate in h^{-1}. The constant γ_j is 2 or can be smaller than 2, depending on the smoothness of the underlying function $b_{n,ij}(x,\tau)$. The value of γ_j does not affect the asymptotic theory of this article, as long as it is strictly positive. In Section 6.4, we provide primitive sufficient conditions to establish the uniform validity of our bootstrap test for the first empirical example.

4. POWER PROPERTIES

In this section, we consider the power properties of the bootstrap test.[13] In Section 4.1, we establish the consistency of our test. Section 4.2 provides heuristic arguments behind local power properties of our tests, and Section 4.3 presents the local power function in a general form.[14]

4.1. Consistency

First, to show consistency of our test, we make the following assumption.

Assumption C1. For each $j \in \mathbb{N}_J$ and $(x,\tau) \in \mathcal{S}$, $v_{n,\tau,j}(x) = v_{\tau,j}(x) + o(1)$, and

$$\limsup_{n\to\infty} \sup_{(x,\tau)\in\mathcal{S}} |v_{n,\tau,j}(x)| < \infty. \tag{4.1}$$

The pointwise convergence $v_{n,\tau,j}(x) = v_{\tau,j}(x) + o(1)$ holds typically by an appropriate choice of $v_{n,\tau,j}(x)$. In many examples, condition (4.1) is often implied by Assumptions A1–A6. If we revisit the simple example considered in Section 2.3, it is straightforward to see that under Assumptions A1–A6, with the subscript τ suppressed, $v_{n,1}(x) = v_1(x) + o(1)$, where $v_{n,1}(x) \equiv \mathbf{E}\hat{v}_{n,1}(x)$ and $v_1(x) \equiv \mathbf{E}(Y|X=x)f(x)$, and (4.1) holds easily.

We now establish the consistency of our proposed test as follows.

THEOREM 3. *Suppose that Assumptions A1–A6, B1–B4, and C1 hold and that we are under a fixed alternative hypothesis such that*

$$\int \Lambda_p\left(v_{\tau,1}(x),\ldots,v_{\tau,J}(x)\right) dQ(x,\tau) > 0.$$

Then as $n \to \infty$,

$$P\{\hat{\theta} > c^*_{a,\eta}\} \to 1.$$

4.2. Local Power Analysis: Definitions and Heuristics

In this section, we investigate the local power properties of our test. For local power analysis, we formally define the space of Pitman directions. Let $\mathcal{D}$

be the collection of $\mathbf{R}^J$-valued bounded functions on $\mathcal{X} \times \mathcal{T}$ such that for each $\delta = (\delta_1, \ldots, \delta_J) \in \mathcal{D}$, $Q\{(x, \tau) \in \mathcal{S} : \delta_j(x, \tau) \neq 0\} > 0$ for some $j = 1, \ldots, J$. That is, at least one of the components of any $\delta \in \mathcal{D}$ is a nonzero function a.e. For each $\delta = (\delta_1, \ldots, \delta_J) \in \mathcal{D}$, we write $\delta_{\tau, j}(x) = \delta_j(x, \tau)$, $j = 1, \ldots, J$.

For a given vector of sequences $b_n = (b_{n,1}, \ldots, b_{n,J})$, such that $b_{n,j} \to \infty$, and $\delta \in \mathcal{D}$, we consider the following type of local alternatives:

$$H_\delta : v_{\tau, j}(x) = v_{\tau, j}^0(x) + \frac{\delta_{\tau, j}(x)}{b_{n, j}}, \text{ for all } j \in \mathbb{N}_J, \tag{4.2}$$

where $v_{\tau, j}^0(x) \leq 0$ for all $(x, \tau, j) \in \mathcal{X} \times \mathcal{T} \times \mathbb{N}_J$, $\delta_{\tau, j}(x) > 0$ for some $(x, \tau, j) \in \mathcal{X} \times \mathcal{T} \times \mathbb{N}_J$ such that $v_{\tau, j}(x) > 0$ for some $(x, \tau, j) \in \mathcal{X} \times \mathcal{T} \times \mathbb{N}_J$. Note that in (4.2), $v_{\tau, j}(x)$ is a sequence of Pitman local alternatives that consist of three components: $v_{\tau, j}^0(x)$, b_n, and $\delta_{\tau, j}(x)$.

The first component $v_{\tau, j}^0(x)$ determines where the sequence of local alternatives converges to. For example, if $v_{\tau, j}^0(x) \equiv 0$ for all (x, τ, j), then we have a sequence of local alternatives that converges to the least favorable case. We allow for negative values for $v_{\tau, j}^0(x)$, so that we include the local alternatives that do not converge to the least favorable case as well.

From here on, we assume the local alternative hypotheses of the form in (4.2). We fix $v_{\tau, j}^0(x)$ and identify each local alternative with a pair (b_n, δ) for each Pitman direction $\delta \in \mathcal{D}$. The following definitions are useful to explain our local power results.

DEFINITION 3.   (i) *Given a Pitman direction $\delta \in \mathcal{D}$, we say that an α-level test, $1\{T > c_\alpha\}$, has nontrivial local power against (b_n, δ), if under the local alternatives (b_n, δ),*

$$\liminf_{n \to \infty} P\{T > c_\alpha\} > \alpha,$$

and say that the test has trivial local power against (b_n, δ), if under the local alternatives (b_n, δ),

$$\limsup_{n \to \infty} P\{T > c_\alpha\} \leq \alpha.$$

(ii) *Given a collection $\mathcal{D}$, we say that a test has convergence rate b_n against $\mathcal{D}$, if the test has nontrivial local power against (b_n, δ) for some $\delta \in \mathcal{D}$, and has trivial local power against (b_n', δ) for all $\delta \in \mathcal{D}$ and all b_n' such that $b_{n,j}'/b_{n,j} \to \infty$ as $n \to \infty$, for all $j = 1, \ldots, J$.*

One of the remarkable aspects of the local power properties is that our test has *two types of convergence rates*. More specifically, there exists a partition $(\mathcal{D}_1, \mathcal{D}_2)$ of $\mathcal{D}$, where our test has a rate b_n against $\mathcal{D}_1$ and another rate b_n' against $\mathcal{D}_2$. Furthermore, in many nonparametric inequality testing environments, the faster of the two rates b_n and b_n' achieves the parametric rate of $\sqrt{n}$.

To see this closely, let us assume the set-up of testing inequality restrictions on a mean regression function in Section 2.4, and consider the following local alternatives:

$$v_{n,1}(x) = v_0(x) + \frac{\delta(x)}{b_n}, \tag{4.3}$$

where $v_0(x) \leq 0$ for all $x \in \mathcal{X}$, and $\delta \in \mathcal{D}$.

First, we set $b_n = \sqrt{n}$. Then under this local alternative hypothesis (b_n, δ), we can verify that with probability approaching one,

$$h^{-1/2}(\hat{\theta} - a_{n,0}) = h^{-1/2} \left\{ \int_{B_n^0(c_n)} \left[Z_{n,1}(x) + \frac{\sqrt{nh} v_0(x)}{\hat{\sigma}_1(x)} + \frac{h^{1/2}\delta(x)}{\hat{\sigma}_1(x)} \right]_+ dx - a_{n,0} \right\}, \tag{4.4}$$

where $Z_{n,1}(x) = \sqrt{nh} \{\hat{v}_1(x) - v_{n,1}(x)\}/\hat{\sigma}_1(x)$, $B_n^0(c_n) = \{x \in \mathcal{X} : |\sqrt{nh} v_0(x)| \leq c_n\}$, $c_n \to \infty$, $\sqrt{\log n}/c_n \to 0$, and

$$a_{n,0} = \mathbf{E} \left[\int_{B_n^0(c_n)} \left[Z_{n,1}(x) \right]_+ dx \right].$$

Under regularity conditions, the right-hand side of (4.4) is approximated by

$$h^{-1/2} \left\{ \int_{B^0(0)} \left[Z_{n,1}(x) + \frac{h^{1/2}\delta(x)}{\sigma_1(x)} \right]_+ dx - a_{n,\delta} \right\} + h^{-1/2} \{a_{n,\delta} - a_{n,0}\}, \tag{4.5}$$

where $B^0(0) = \{x \in \mathcal{X} : v_0(x) = 0\}$ and

$$a_{n,\delta} = \mathbf{E} \left[\int_{B^0(0)} \left[Z_{n,1}(x) + \frac{h^{1/2}\delta(x)}{\sigma_1(x)} \right]_+ dx \right].$$

The leading term in (4.5) converges in distribution to $\mathbb{Z}_1 \sim N(0, \sigma_0^2)$ precisely as in (2.10). Furthermore, we can show that

$$a_{n,\delta} = \int_{B^0(0)} \mathbf{E} \left[\mathbb{Z}_1 + \frac{h^{1/2}\delta(x)}{\sigma_1(x)} \right]_+ dx + o(h^{1/2}) \text{ and}$$

$$a_{n,0} = \int_{B^0(0)} \mathbf{E}[\mathbb{Z}_1]_+ dx + o(h^{1/2}).$$

Therefore, as for the last term in (4.5), we find that

$$h^{-1/2} \{a_{n,\delta} - a_{n,0}\} = \int_{B^0(0)} h^{-1/2} \left(\mathbf{E} \left[\mathbb{Z}_1 + \frac{h^{1/2}\delta(x)}{\sigma_1(x)} \right]_+ - \mathbf{E}[\mathbb{Z}_1]_+ \right) dx + o(1)$$

$$= 2\phi(0) \int_{B^0(0)} \frac{\delta(x)}{\sigma_1(x)} dx + o(1),$$

where the last equality follows from expanding $h^{-1/2}\left\{\mathbf{E}\left[\mathbb{Z}_1+h^{1/2}\delta(x)/\sigma_1(x)\right]_+ - \mathbf{E}[\mathbb{Z}_1]_+\right\}$. We conclude that under the local alternatives, we have

$$h^{-1/2}(\hat{\theta}-a_{n,0}) \to_d \mathbb{Z}_1 + 2\phi(0)\int_{B^0(0)}\frac{\delta(x)}{\sigma_1(x)}dx.$$

The magnitude of the last term in the limit determines the local power of the test. Thus under Pitman local alternatives such that

$$\int_{B^0(0)}\frac{\delta(x)}{\sigma_1(x)}dx > 0, \tag{4.6}$$

the test has nontrivial power against $\sqrt{n}$-converging Pitman local alternatives. Note that the integral in (4.6) is defined on the population contact set $B^0(0)$. Thus, the test has nontrivial power, unless the contact set has Lebesgue measure zero or $\delta(\cdot)$ is "too often negative" on the contact set.

When the integral in (4.6) is zero, we consider the local alternatives (b_n, δ) with a slower convergence rate $b_n = n^{1/2}h^{1/4}$. Following similar arguments as before, we now have

$$h^{-1/2}(\hat{\theta}-a_{n,0}) \to_d \mathbb{Z}_1 + \lim_{n\to\infty}h^{-1/2}\left\{\bar{a}_{n,\delta}-a_{n,0}\right\},$$

where

$$\bar{a}_{n,\delta} = \int_{B^0(0)}\mathbf{E}\left[Z_{n,1}(x)+\frac{h^{1/4}\delta(x)}{\sigma_1(x)}\right]_+ dx,$$

which can be shown again to be equal to

$$\int_{B^0(0)}\mathbf{E}\left[\mathbb{Z}_1+\frac{h^{1/4}\delta(x)}{\sigma_1(x)}\right]_+ dx + o(h^{1/2}).$$

However, observe that

$$h^{-1/2}\int_{B^0(0)}\left\{\mathbf{E}\left[\mathbb{Z}_1+\frac{h^{1/4}\delta(x)}{\sigma_1(x)}\right]_+ - \mathbf{E}[\mathbb{Z}_1]_+\right\}dx$$

$$= h^{-1/4}2\phi(0)\int_{B^0(0)}\frac{\delta(x)}{\sigma_1(x)}dx + \frac{1}{2}\int_{B^0(0)}\frac{\delta^2(x)}{\sigma_1^2(x)}dx + o(1)$$

$$= \frac{1}{2}\int_{B^0(0)}\frac{\delta^2(x)}{\sigma_1^2(x)}dx + o(1)$$

because $\int_{B^0(0)}\{\delta(x)/\sigma_1(x)\}dx = 0$. We find that under the local alternative hypothesis in (4.3) with $b_n = n^{1/2}h^{1/4}$,

$$h^{-1/2}(\hat{\theta}-a_{n,0}) \to_d \mathbb{Z}_1 + \frac{1}{2}\int_{B^0(0)}\frac{\delta^2(x)}{\sigma_1^2(x)}dx.$$

Therefore, even when $\int_{B^0(0)} \{\delta(x)/\sigma_1(x)\}dx = 0$, the test still has nontrivial power against $n^{1/2}h^{1/4}$-converging Pitman local alternatives, if the Pitman directions are such that

$$\int_{B^0(0)} \{\delta^2(x)/\sigma_1^2(x)\}dx > 0.$$

Now let us consider the partition $(\mathcal{D}_1, \mathcal{D}_2)$ of $\mathcal{D}$, where,

$$\mathcal{D}_1 = \left\{ \delta \in \mathcal{D} : \int_{B^0(0)} \delta(x)/\sigma_1(x)dx \neq 0 \right\} \text{ and}$$

$$\mathcal{D}_2 = \left\{ \delta \in \mathcal{D} : \int_{B^0(0)} \delta(x)/\sigma_1(x)dx = 0 \text{ and } \int_{B^0(0)} \{\delta^2(x)/\sigma_1^2(x)\}dx > 0 \right\}.$$

When $\inf_{x \in \mathcal{X}} \sigma_1^2(x) > c > 0$ for some $c > 0$ (recall Assumption A5) and $Q(B^0(0)) > 0$, we have $\int_{B^0(0)} \{\delta^2(x)/\sigma_1^2(x)\}dx > 0$ and the set $\{\mathcal{D}_1, \mathcal{D}_2\}$ becomes a partition of $\mathcal{D}$. Thus the bootstrap test has a convergence rate of $\sqrt{n}$ against $\mathcal{D}_1$ and $n^{1/2}h^{1/4}$-rate against $\mathcal{D}_2$. In the next section, Corollary 1 provides a general result of this phenomenon of dual convergence rates of our bootstrap test.

4.3. Local Power Analysis: Results

We now provide general local power functions explicitly. We first present explicit forms of location and scale normalizers, $a_n(q_n)$ and $\sigma_n(q_n)$ in (3.9). Let for $j, k \in \mathbb{N}_J$, and $\tau_1, \tau_2 \in \mathcal{T}$,

$$\rho_{n,\tau_1,\tau_2,j,k}(x,u) \equiv \frac{1}{h^d} \mathbf{E} \left[\beta_{n,x,\tau_1,j} \left(Y_{ij}, \frac{X_i - x}{h} \right) \beta_{n,x,\tau_2,k} \left(Y_{ik}, \frac{X_i - x}{h} + u \right) \right]. \quad (4.7)$$

This function approximates the asymptotic covariance between $\sqrt{n}(\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x))/\hat{\sigma}_{\tau,j}(x)$ and $\sqrt{n}(\hat{v}_{\tau,j}(x+uh) - v_{n,\tau,j}(x+uh))/\hat{\sigma}_{\tau,j}(x)$. We define $\Sigma_{n,\tau_1,\tau_2}(x,u)$ to be the J-dimensional square matrix with (j,k)-th entry given by $\rho_{n,\tau_1,\tau_2,j,k}(x,u)$.

Define for $\mathbf{v} \in \mathbf{R}^J$,

$$\bar{\Lambda}_{x,\tau}(\mathbf{v}) \equiv \sum_{A \in \mathcal{N}_J} \Lambda_{A,p}(\mathbf{v}) 1\left\{ (x,\tau) \in B_{n,A}(q_n) \right\}.$$

Then we define

$$a_n(q_n) \equiv \int_{\mathcal{X}} \int_{\mathcal{T}} \mathbf{E}\left[\bar{\Lambda}_{x,\tau_1}(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0)) \right] d\tau dx,$$

and

$$\sigma_n^2(q_n) \equiv \int_{\mathcal{U}} \int_{\mathcal{X}} \int_{\mathcal{T}} \int_{\mathcal{T}} C_{n,\tau_1,\tau_2}(x,u) d\tau_1 d\tau_2 dx du, \quad (4.8)$$

where

$$C_{n,\tau_1,\tau_2}(x,u) \equiv Cov\left(\bar{\Lambda}_{x,\tau_1}(\mathbb{W}_{n,\tau_1,\tau_2}^{(1)}(x,u)), \bar{\Lambda}_{x,\tau_2}(\mathbb{W}_{n,\tau_1,\tau_2}^{(2)}(x,u)) \right),$$

and $[\mathbb{W}_{n,\tau_1,\tau_2}^{(1)}(x,u)^\top, \mathbb{W}_{n,\tau_1,\tau_2}^{(2)}(x,u)^\top]^\top$ is a mean zero $\mathbf{R}^{2J}$-valued Gaussian random vector whose covariance matrix is given by

$$\begin{bmatrix} \Sigma_{n,\tau_1,\tau_1}(x,0) & \Sigma_{n,\tau_1,\tau_2}(x,u) \\ \Sigma_{n,\tau_1,\tau_2}(x,u)^\top & \Sigma_{n,\tau_2,\tau_2}(x+uh,0) \end{bmatrix}. \tag{4.9}$$

The multiple integral in (4.8) is nonnegative. The limit of the quantity $\sigma_n^2(q_n)$ as $n \to \infty$, if it is positive, is nothing but the asymptotic variance of the test statistic $\hat\theta$ (after location-scale normalization).

We first make the following assumptions.

Assumption C2. (i) For each $(\tau,j) \in \mathcal{T} \times \mathbb{N}_J$, there exists a map $v_{n,\tau,j}^0 : \mathbf{R}^d \to \mathbf{R}$ such that for each $x \in \mathcal{S}_\tau(\varepsilon_1)$, $v_{n,\tau,j}^0(x) \le 0$, and

$$v_{n,\tau,j}(x) = v_{n,\tau,j}^0(x) + \frac{\delta_{\tau,j}(x)}{b_{n,j}}(1+o(1)), \tag{4.10}$$

where $o(1)$ is uniform in $x \in \mathcal{S}_\tau$ and in $\tau \in \mathcal{T}$, as $n \to \infty$ and $b_{n,j} \to \infty$ is the positive sequence in (4.2).
(ii) $\sup_{(x,\tau)\in\mathcal{S}} |\sigma_{n,\tau,j}(x) - \sigma_{\tau,j}(x)| = o(1)$, as $n \to \infty$, for some function $\sigma_{\tau,j}(x)$ such that $\inf_{(x,\tau)\in\mathcal{S}} \sigma_{\tau,j}(x) > 0$.

Assumption C2 can also be shown to hold in many examples. When appropriate smoothness conditions for $v_{\tau,j}(x)$ hold and a suitable (possibly higher-order) kernel function is used, we can take $v_{n,\tau,j}(x)$ in Assumption A1 to be identical to $v_{\tau,j}(x)$, and hence Assumption C2 is implied by (4.2). For the simple example in Section 2.4, if we take $v_{n,j}(x) = \mathbf{E}\hat{v}_j(x)$, it follows that $v_{n,j}(x) = v_{n,j}^0(x) + b_{n,j}^{-1} \int \delta_j(x+zh)K(z)dz$, with $v_{n,j}^0(x) = \int v_j^0(x+zh)K(z)dz$. Hence when $\delta_j(x)$ is uniformly continuous in x, we obtain Assumption C2.

The local asymptotic power function is based on the asymptotic normal approximation of the distribution of $\hat\theta$ (after scale and location normalization) under the local alternatives. For this purpose, we define the sequence of probability sets that admit the normal approximation under local alternatives. For $c_1, c_2 > 0$, let $B_n^0(c_1,c_2)$ and $B_{n,A}^0(c_1,c_2)$ denote $B_n(c_1,c_2)$ and $B_{n,A}(c_1,c_2)$ except that $v_{n,\tau,j}(x)$'s are replaced by $v_{n,\tau,j}^0(x)$'s in Assumption C2. As before, we write $B_n^0(c) \equiv B_n^0(c,c)$.

DEFINITION 4. *For any positive sequence $\lambda_n \to 0$, define*

$$\mathcal{P}_n^0(\lambda_n) \equiv \left\{ P \in \tilde{\mathcal{P}}_n^0(\lambda_n) : \sigma_n^2(0) \ge \eta/\Phi^{-1}(1-\alpha) \right\},$$

where $\tilde{\mathcal{P}}_n^0(\lambda_n)$ is equal to $\tilde{\mathcal{P}}_n(\lambda_n, q_n)$ except that $B_{n,A}(c_{n,U}, c_{n,L})$ and $B_{n,A}(q_n)$ are replaced by $B_{n,A}^0(c_{n,U}, c_{n,L})$ and $B_{n,A}^0(q_n)$ for all $A \in \mathcal{N}_J$, and q_n is set to be zero.

To give a general form of the local power function, let us define $\psi_{n,A,\tau}(\cdot;x) : \mathbf{R}^J \to [0,\infty)$, $(x,\tau) \in \mathcal{X} \times \mathcal{T}$ and $A \subset \mathbb{N}_J$, as

$$\psi_{n,A,\tau}(\mathbf{y};x) = \frac{1}{\sigma_n(0)}\mathbf{E}\left[\Lambda_{A,p}\left(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0)+\mathbf{y}\right)\right] \cdot 1\left\{(x,\tau) \in B_{n,A}^0(0)\right\}.$$

The local power properties of the bootstrap test are mainly determined by the slope and the curvature of this function. So, we define

$$\psi_{n,A,\tau}^{(1)}(\mathbf{y};x) \equiv \frac{\partial}{\partial\mathbf{y}}\psi_{n,A,\tau}(\mathbf{y};x) \text{ and } \psi_{n,A,\tau}^{(2)}(\mathbf{y};x) \equiv \frac{\partial^2}{\partial\mathbf{y}\partial\mathbf{y}^\top}\psi_{n,A,\tau}(\mathbf{y};x), \quad \textbf{(4.11)}$$

if the first derivatives and the second derivatives in the definition exist, respectively.

Assumption C3.   (i) There exists $\varepsilon_1 > 0$ such that for all $(\tau, A) \in \mathcal{T} \times \mathcal{N}_J$ and all x in the interior of $\mathcal{S}_\tau(\varepsilon_1)$, $\psi_{n,A,\tau}^{(1)}(\mathbf{0};x)$ exists for all $n \geq 1$ and

$$\psi_{A,\tau}^{(1)}(\mathbf{0};x) \equiv \lim_{n\to\infty}\psi_{n,A,\tau}^{(1)}(\mathbf{0};x)$$

exists, and $\limsup_{n\to\infty}\sup_{(x,\tau)\in\mathcal{S}}|\psi_{n,A,\tau}^{(1)}(\mathbf{0};x)| < C$ *for some $C > 0$.*

(ii) There exists $\varepsilon_1 > 0$ such that for all $(\tau, A) \in \mathcal{T} \times \mathcal{N}_J$ and all x in the interior of $\mathcal{S}_\tau(\varepsilon_1)$, $\psi_{n,A,\tau}^{(2)}(\mathbf{0};x)$ exists for all $n \geq 1$ and

$$\psi_{A,\tau}^{(2)}(\mathbf{0};x) \equiv \lim_{n\to\infty}\psi_{n,A,\tau}^{(2)}(\mathbf{0};x)$$

exists, and $\limsup_{n\to\infty}\sup_{(x,\tau)\in\mathcal{S}}|\psi_{n,A,\tau}^{(2)}(\mathbf{0};x)| < C$ for some $C > 0$.

To appreciate Assumption C3, consider the case where $J = 2$, $A = \{1, 2\}$, and $\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0)$ has a distribution denoted by G_n. Choose $y_1 \geq y_2$ without loss of generality. We take $\Lambda_p(v_1, v_2) = \max\{v_1, v_2, 0\}^p$. Then we can write $\mathbf{E}[\Lambda_{A,p}(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0)+\mathbf{y})]$ as

$$\int_{\mathbf{R}^2}(w_1+y_1)^p 1\{w_1 \in [w_2+y_2-y_1, \infty) \text{ and } w_2 \in [-y_2, \infty)\}dG_n(w_1, w_2)$$

$$+ \int_{\mathbf{R}^2}(w_2+y_2)^p 1\{w_1 \in (-\infty, w_2+y_2-y_1) \text{ and } w_2 \in [-y_2, \infty)\}dG_n(w_1, w_2)$$

$$+ \int_{\mathbf{R}^2}(w_1+y_1)^p 1\{w_1 \in [-y_1, \infty) \text{ and } w_2 \in (-\infty, -y_2)\}dG_n(w_1, w_2).$$

Certainly the three quantities are all differentiable in (y_1, y_2).

The following theorem offers the local power function of the bootstrap test in a general form.

THEOREM 4. *Suppose that Assumptions A1–A6, B1–B4, C1–C2, and C3(i) hold and that*

$$h^{-d/2}(\log n)^{p/2}\lambda_n \to 0, \quad \textbf{(4.12)}$$

as $n \to \infty$. Then for each sequence $P_n \in \mathcal{P}_n^0(\lambda_n)$, $n \geq 1$, which satisfies the local alternative hypothesis (b_n, δ) for some $\delta \in \mathcal{D}$ with $b_n = (r_{n,j} h^{-d/2})_{j=1}^J$,

$$\lim_{n \to \infty} P_n\{\hat{\theta} > c_{\alpha,\eta}^*\} = 1 - \Phi(z_{1-\alpha} - \mu_1(\delta)),$$

where Φ denotes the standard normal cdf,

$$\mu_1(\delta) \equiv \sum_{A \in \mathcal{N}_J} \int \psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma}(x) \, dQ(x,\tau),$$

and

$$\delta_{\tau,\sigma}(x) \equiv \left(\frac{\delta_{\tau,1}(x)}{\sigma_{\tau,1}(x)}, \ldots, \frac{\delta_{\tau,J}(x)}{\sigma_{\tau,J}(x)} \right). \tag{4.13}$$

Theorem 4 shows that if we take b_n such that $b_{n,j} = r_{n,j} h^{-d/2}$ for each $j = 1, \ldots, J$, the local asymptotic power of the test against (b_n, δ) is determined by the shift $\mu_1(\delta)$. Thus, the bootstrap test has nontrivial local power against (b_n, δ) if and only if

$$\mu_1(\delta) > 0.$$

The test is asymptotically biased against (b_n, δ) such that $\mu_1(\delta) < 0$.

Suppose that

$$\mu_1(\delta) = 0, \tag{4.14}$$

for all $A \in \mathcal{N}_J$, i.e., when $\delta_{\tau,\sigma}$ has positive and negative parts which precisely cancels out in the integration. Then, we show that the bootstrap test has nontrivial asymptotic power against local alternatives that converges at a rate slower than $n^{-1/2}$ to the null hypothesis.

THEOREM 5. *Suppose that the conditions of Theorem 4 and Assumption C3(ii) hold. Then for each sequence* $P_n \in \mathcal{P}_n^0(\lambda_n)$, $n \geq 1$, *which satisfies the local alternative hypothesis* (b_n, δ) *for some* $\delta \in \mathcal{D}$ *such that* $\mu_1(\delta) = 0$ *and* $b_n = (r_{n,j} h^{-d/4})_{j=1}^J$,

$$\lim_{n \to \infty} P_n\{\hat{\theta} > c_{\alpha,\eta}^*\} = 1 - \Phi(z_{1-\alpha} - \mu_2(\delta)),$$

where

$$\mu_2(\delta) \equiv \frac{1}{2} \sum_{A \in \mathcal{N}_J} \int \delta_{\tau,\sigma}^\top(x) \psi_{A,\tau}^{(2)}(\mathbf{0};x) \delta_{\tau,\sigma}(x) \, dQ(x,\tau).$$

The local power function depends on the limit of the curvature of the function $\psi_{n,A,\tau}(\mathbf{y};x)$ at $\mathbf{y} = \mathbf{0}$, for all $A \in \mathcal{N}_J$. When the function is strictly concave

at $\mathbf{0}$ in the limit, $\psi_{A,\tau}^{(2)}(\mathbf{0}; x)$ is positive definite on $\mathcal{X} \times \mathcal{T}$, and in this case, the bootstrap test has nontrivial power whenever $\delta_{\tau,\sigma}(x)$ is nonzero on a set whose intersection with $B_n^0(0)$ has Lebesgue measure greater than $c > 0$ for all $n \geq 1$, for some $c > 0$.

From Theorems 4 and 5, it is seen that the phenomenon of dual convergence rates generally holds for our tests. To formally state the result, define

$$\mathcal{D}_1 \equiv \{\delta \in \mathcal{D} : \mu_1(\delta) \neq 0\} \text{ and}$$
$$\mathcal{D}_2 \equiv \{\delta \in \mathcal{D} : \mu_1(\delta) = 0 \text{ and } \mu_2(\delta) > 0\}.$$

When $\liminf_{n\to\infty} Q(B_n^0(0)) > 0$, the set $\{\mathcal{D}_1, \mathcal{D}_2\}$ becomes a partition of the space of Pitman directions $\mathcal{D}$.

COROLLARY 1. *Suppose that the conditions of Theorem 5 hold. Then the bootstrap test has convergence rate $b_n = (r_{n,j}h^{-d/2})_{j=1}^J$ against $\mathcal{D}_1$, and convergence rate $b_n = (r_{n,j}h^{-d/4})_{j=1}^J$ against $\mathcal{D}_2$.*

When $r_{n,j}$'s diverge to infinity at the usual nonparametric rate $r_{n,j} = n^{1/2}h^{d/2}$ as in many kernel-based estimators, the test has a parametric rate of convergence $b_n = \sqrt{n}$ and nontrivial local power against $\mathcal{D}_1$. However, the test has a convergence rate slower than the parametric rate against $\mathcal{D}_2$.

5. MONTE CARLO EXPERIMENTS

In this section, we report the finite-sample performance of our proposed test for the Monte Carlo design considered in (Andrews and Shi, 2013, Sect. 10.3, hereafter AS). The null hypothesis has the form

$$H_0 : \mathbf{E}(Y - \theta | X = x) \leq 0 \text{ for each } x \in \mathcal{X}$$

with a fixed θ. AS generated a random sample of (Y, X) from the following model:

$$Y = f(X) + U,$$

where $X \sim \text{Unif}[-2, 2]$, U follows truncated normal such that $U_i = \min\{\max\{-3, \sigma\tilde{U}_i\}, 3\}$ with $\tilde{U}_i \sim N(0, 1)$ and $\sigma = 1$, and $f(\cdot)$ is a function with an alternative shape. AS considered two functions:

$$f_{AS1}(x) := L\phi(x^{10}),$$
$$f_{AS2}(x) := L \cdot \max\{\phi((x - 1.5)^{10}), \phi((x + 1.5)^{10})\}.$$

These two functions have steep slopes, f_{AS1} being a roughly plateau-shaped function and f_{AS2} a roughly double-plateau-shaped function, respectively. AS considered the following Monte Carlo designs:

DGP1: $f(x) = f_{AS1}(x)$ and $L = 1$;   DGP2: $f(x) = f_{AS1}(x)$ and $L = 5$;
DGP3: $f(x) = f_{AS2}(x)$ and $L = 1$;   DGP4: $f(x) = f_{AS2}(x)$ and $L = 5$.

AS compared their tests with (Chernozhukov et al., 2013, hereafter CLR) and Lee et al. (2013). The latter test uses conservative standard normal critical values based on the least favorable configuration.

In this article, we used the same statistic for Lee et al. (2013) as reported in AS. Specifically, we used the L_1 version of the test with the inverse standard error weight function. In implementing the test, we used $K(u) = (3/2)(1 - (2u)^2)I(|u| \leq 1/2)$ and $h = 2 \times \hat{s}_X \times n^{-1/5}$, where $I(A)$ is the usual indicator function that has value one if A is true and zero otherwise and $\hat{s}_X$ is the sample standard deviation of X. Thus, the only difference between the new test (which we call LSW2) and Lee et al. (2013) (which we call LSW1) is the use of critical values: LSW1 uses the standard normal critical values based on the least favorable configuration, whereas LSW2 uses bootstrap critical values based on the estimated contact set. See the next subsection for details regarding contact set estimation.

The experiments considered sample sizes of $n = 100, 250, 500, 1,000$, and the nominal level of $\alpha = 0.05$. We performed 1,000 Monte Carlo replications in each experiment. The number of bootstrap replications was 200.

The null hypothesis is tested on $\mathcal{X} = [-1.8, 1.8]$. To compare simulation results from AS, the coverage probability (CP) is computed at nominal level 95% when $\theta = \max_{x \in \mathcal{X}} f(x)$ and the false coverage probability (FCP) is computed at nominal level 95% when $\theta = \max_{x \in \mathcal{X}} f(x) - 0.02$.

5.1. Obtaining $\hat{c}_n$

To construct $\hat{c}_n$, we suggest the following procedure. First, define

$$S_n^* \equiv \max \left\{ \sup_{(j,\tau,x)} \hat{s}_{\tau,j}^*(x), \sqrt{\log n} \right\}.$$

Then, set

$$\hat{c}_n = C_{cs}(\log \log n) q_n(S_n^*), \tag{5.1}$$

where $q_n(S_n^*)$ is the $(1 - 0.1/\log n)$ quantile of the bootstrap distribution of S_n^*, and $C_{cs} \in \{0.4, 0.5, 0.6\}$.

Although the rule-of-thumb for $\hat{c}_n$ in (5.1) is not completely data-driven, it has the advantage that even when $\hat{u}_{\tau,j}(x)$ is not scale-invariant (say, by choosing $\hat{\sigma}_{\tau,j}(x) \equiv 1$), the tuning parameter $\hat{c}_n$ automatically adjusts to any change in the scale of $\hat{u}_{\tau,j}(x)$, due to the term $q_n(S_n^*)$; see Chernozhukov et al. (2013) for a similar idea.[15] This data-dependent choice of $\hat{c}_n$ is encompassed by the theoretical framework of this article, while many other choices are also admitted.[16] Lastly, recall that $\eta \equiv 10^{-3}$ is a small fixed number.

5.2. Simulation Results

Tables 1 and 2 report the results of Monte Carlo experiments. In each table, figures in columns (1)–(5) are from Table V of Andrews and Shi (2013), whereas those

TABLE 1. Results for Monte Carlo experiments: coverage probability

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		AS		CLR		LSW1		LSW2	
	n	CvM	KS	series	local linear		$C_{cs} = 0.4$	$C_{cs} = 0.5$	$C_{cs} = 0.6$
DGP1	100	0.986	0.986	0.707	0.804	1.00	0.982	0.994	0.998
	250	0.975	0.973	0.805	0.893	1.00	0.971	0.976	0.979
	500	0.975	0.970	0.872	0.925	1.00	0.953	0.963	0.974
	1,000	0.971	0.966	0.909	0.935	1.00	0.964	0.971	0.975
DGP2	100	1.00	1.00	0.394	0.713	0.994	1.00	1.00	
	250	1.00	1.00	0.683	0.856	1.00	0.969	0.976	0.986
	500	1.00	1.00	0.833	0.908	1.00	0.953	0.959	0.966
	1,000	1.00	1.00	0.900	0.927	1.00	0.965	0.970	0.973
DGP3	100	0.970	0.969	0.620	0.721	0.973	0.987	0.995	
	250	0.969	0.964	0.762	0.854	1.00	0.963	0.972	0.976
	500	0.963	0.957	0.854	0.900	1.00	0.955	0.962	0.966
	1,000	0.969	0.963	0.901	0.927	1.00	0.948	0.954	0.957
DGP4	100	0.998	0.999	0.321	0.655	1.00	0.992	1.00	1.00
	250	0.997	0.998	0.612	0.826	1.00	0.966	0.973	0.983
	500	0.994	0.994	0.808	0.890	1.00	0.955	0.957	0.962
	1,000	0.994	0.991	0.893	0.918	1.00	0.950	0.956	0.958

Notes: Figures in columns (1)–(5) are from Table V of Andrews and Shi (2013), whereas those in columns (6)–(8) are based 1,000 Monte Carlo replications in each experiment, with the number of bootstrap replications being 200. LSW1 refers to the test of Lee et al. (2013), which uses conservative standard normal critical values based on the least favorable configuration. LSW2 refers to this paper that uses bootstrap critical values based on the estimated contact set. The tuning parameter is chosen by the rule $\hat{c}_n = C_{cs} \log\log(n) q_{1-0.1/\log(n)}(S_n^*)$, where $C_{cs} \in \{0.4, 0.5, 0.6\}$.

in columns (6)–(8) are from our Monte Carlo experiments. Table 1 shows that coverage probabilities of LSW2 are much closer to the nominal level than those of LSW1. When $C_{cs} = 0.4$ and $n = 100$ or 250, we see some under-coverage for LSW2, but it disappears as n gets larger. Table 2 reports the false coverage probabilities (FCPs). Figures in columns (1)–(5) are "CP-corrected" by AS, where those in columns (6)–(8) are not "CP-corrected". However, CP-correction would not change the results for either $n \geq 500$ or $c \geq 0.5$ since in each of these cases, we have over-coverage. We can see that in terms of FCPs, LSW2 performs much better than LSW1 in all DGPs. Furthermore, the performance of LSW2 is equivalent to that of AS for DGP1, DGP3, and DGP4, and is superior to AS for DGP2. Overall, our simulation results show that our new test is a substantially improved version of LSW1 and is now very much comparable to AS. The relatively poor performance of CLR in Tables 1 and 2 are mainly due to the experimental design. If the underlying function is sharply peaked, as those in the reported simulations of Chernozhukov et al. (2013), CLR performs better than AS. In unreported simulations, we confirmed that CLR performs better than LSW2 as well. This is very

TABLE 2. Results for Monte Carlo experiments: false coverage probability

	n	(1) CvM	(2) KS	(3) series	(4) local linear	(5) LSW1	(6) $C_{cs} = 0.4$	(7) $C_{cs} = 0.5$	(8) $C_{cs} = 0.6$
		AS		CLR				LSW2	
DGP1	100	0.84	0.89	0.88	0.83	0.98	0.85	0.92	0.97
	250	0.57	0.67	0.82	0.69	0.92	0.43	0.48	0.53
	500	0.25	0.37	0.72	0.50	0.70	0.18	0.19	0.20
	1,000	0.03	0.07	0.57	0.26	0.25	0.02	0.02	0.02
DGP2	100	1.0	1.0	0.91	0.89	0.99	0.97	0.99	1.0
	250	1.0	1.0	0.85	0.73	0.96	0.47	0.53	0.62
	500	0.97	0.99	0.77	0.56	0.82	0.20	0.23	0.25
	1,000	0.70	0.89	0.61	0.33	0.40	0.03	0.03	0.03
DGP3	100	0.70	0.79	0.89	0.84	0.90	0.69	0.78	0.87
	250	0.30	0.46	0.83	0.66	0.65	0.28	0.32	0.35
	500	0.06	0.15	0.70	0.47	0.26	0.05	0.06	0.06
	1,000	0.00	0.01	0.55	0.23	0.02	0.00	0.00	0.00
DGP4	100	0.95	0.99	0.91	0.88	0.95	0.90	0.96	0.98
	250	0.66	0.83	0.86	0.70	0.75	0.31	0.36	0.42
	500	0.23	0.42	0.74	0.51	0.36	0.06	0.06	0.07
	1,000	0.01	0.04	0.59	0.29	0.04	0.00	0.00	0.00

Notes: See notes in Table 1. Figures in columns (1)–(5) are "CP-corrected", where those in columns (6)–(8) are not "CP-corrected".

reasonable since CLR is based on the sup-norm statistic, whereas ours is based on the one-sided L_p norm. Therefore, we may conclude that AS, CLR, and LSW2 complement each other.

6. EMPIRICAL EXAMPLE 1: TESTING FUNCTIONAL INEQUALITIES IN AUCTION MODELS

In this example, we go back to the auction environment of GPV mentioned earlier. We first state the testing problem formally, give the form of test statistic, and present empirical results.

6.1. Testing Problem

Suppose that the number I of bidders can take two values, 2 and 3 (that is, $I \in \{2, 3\}$). For each τ such that $0 < \tau < 1$, let $q_k(\tau | x)$ denote the τ-th conditional quantile (given $X = x$) of the observed equilibrium bid distribution when the number of bidders is $I = k$, where $k = 2, 3$. A conditional version of equation (5) of GPV (with $I_1 = 2$ and $I_2 = 3$ in their notation) provides the following

testing restrictions:

$$q_2(\tau|x) - q_3(\tau|x) < 0, \text{ and}$$
$$\underline{b} - 2q_2(\tau|x) + q_3(\tau|x) < 0,$$

(6.1)

for any $\tau \in (0,1]$ and for any $x \in \text{supp}(X)$, where $\text{supp}(X)$ is the (common) support of X, and $\underline{b}$ is the left endpoint of the support of the observed bids.[17] The restrictions in (6.1) are based on conditionally exogenous participation for which the latent private value distribution is independent of the number of bidders conditional on observed characteristics (X), e.g., appraisal values. A slightly weaker version of (6.1) can be put into our general problem of testing the null hypothesis:[18]

$$v_{\tau,1}(x) \equiv q_2(\tau|x) - q_3(\tau|x) \le 0, \text{ and}$$
$$v_{\tau,2}(x) \equiv \underline{b} - 2q_2(\tau|x) + q_3(\tau|x) \le 0,$$

(6.2)

for any $(\tau, x) \in \mathcal{T} \times \mathcal{X} \subset (0,1] \times \text{supp}(X)$.

The example in (6.2) illustrates that in order to test the implications of auction theory, it is essential to test the null hypothesis uniformly in τ and x. More specifically, testing for a wide range of τ is important because testable implications are expressed in terms of conditional stochastic dominance relations. Furthermore, testing the relations uniformly over x is natural since theoretical predictions given by conditionally exogenous participation should hold for any realization of observed auction heterogeneity. It also shows that it is important to go beyond the $J = 1$ case and to include a general $J > 1$. In fact, if the number of bidders can take more than two values, there could be many more functional inequalities (see Corollary 1 of GPV). Finally, we note that $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$ are not forms of conditional moment inequalities and each involves two different conditional quantile functions indexed by τ. Therefore, tests developed for conditional moment inequalities are not directly applicable to this empirical example. There exist related but distinct article regarding this empirical example. See, e.g., Marmer, Shneyerov, and Xu (2013) who developed a nonparametric test for selective entry, and Gimenes and Guerre (2013) who proposed augmented quantile regression for first-price auction models.

6.2. Test Statistic

To implement the test, it is necessary to estimate conditional quantile functions. In estimation of $q_j(\tau|x)$, $j = 2, 3$, we may use a local polynomial quantile regression estimator, say $\widehat{q}_j(\tau|x)$. Now write

$$\hat{v}_{\tau,1}(x) = \hat{q}_2(\tau|x) - \hat{q}_3(\tau|x),$$
$$\hat{v}_{\tau,2}(x) = \underline{\hat{b}} - 2\hat{q}_2(\tau|x) + \hat{q}_3(\tau|x),$$

where $\underline{\hat{b}}$ is a consistent estimator of $\underline{b}$.[19] Then testing (6.2) can be carried out using $\{\hat{v}_{\tau,j}(x) : j = 1, 2\}$ based on our general framework. In this application, our test statistics take the following forms:

$$\hat{\theta}_{\text{sum}} = \int_{\mathcal{X} \times \mathcal{T}} [r_n \hat{v}_{\tau,1}(x)]_+^p \, dQ(x, \tau) + \int_{\mathcal{T} \times \mathcal{X}} [r_n \hat{v}_{\tau,2}(x)]_+^p \, dQ(x, \tau), \text{ or}$$

$$\hat{\theta}_{\text{max}} = \int_{\mathcal{X} \times \mathcal{T}} \left(\max \left\{ [r_n \hat{v}_{\tau,1}(x)]_+, [r_n \hat{v}_{\tau,2}(x)]_+ \right\} \right)^p dQ(x, \tau),$$

(6.3)

where $r_n = n^{1/2} h^{d/2}$. Note that in (6.3), we set $\hat{\sigma}_{\tau, j}(x) \equiv 1$.

6.3. Details on Estimating Conditional Quantile Functions

Assume that $q_k(\tau | x)$ is $(r + 1)$-times continuously differentiable with respect to x, where $r \geq 1$. We use a local polynomial estimator $\widehat{q}_k(\tau | x)$. For $u \equiv (u_1, \ldots, u_d)$, a d-dimensional vector of nonnegative integers, let $[u] = u_1 + \cdots + u_d$. Let A_r be the set of all d-dimensional vectors u such that $[u] \leq r$, and let $|A_r|$ denote the number of elements in A_r. For $z = (z_1, \ldots, z_d)^\top \in \mathbf{R}^d$ with $u = (u_1, \ldots, u_d)^\top \in A_r$, let $z^u = \prod_{m=1}^d z_m^{u_m}$. Now define $c(z) = (z^u)_{u \in A_r}$, for $z \in \mathbf{R}^d$. Note that $c(z)$ is a vector of dimension $|A_r|$.

Let $\{(B_{\ell i}, X_i, L_i) : \ell = 1, \ldots, L_i, i = 1, \ldots, n\}$ denote the observed data, where $\{B_{\ell i} : \ell = 1, \ldots, L_i\}$ denotes the L_i number of observed bids in the i-th auction, X_i a vector of observed characteristics for the i-th auction, and L_i the number of bids for the i-th auction, taking values from $\mathbb{N}_L \equiv \{2, \ldots, \bar{L}\}$. In our application, $\bar{L} = 3$.

Assume that the data $\{(B_{\ell i}, X_i, L_i) : \ell = 1, \ldots, L_i, i = 1, \ldots, n\}$ are i.i.d. over i and that $B_{\ell i}$'s are also i.i.d. over ℓ conditional on X_i and L_i. To implement the test, it is necessary to estimate $\underline{b}$. In our application, we use $\underline{\hat{b}} = \min\{B_{\ell i} : \ell = 1, \ldots, L_i, i = 1, \ldots, n\}$, that is, the overall sample minimum.

For each $x = (x_1, \ldots, x_d)$, the r-th order local polynomial quantile regression estimator of $q_k(\tau | x)$ can be obtained by minimizing

$$S_{n,x,\tau,k}(\gamma) \equiv \sum_{i=1}^n 1\{L_i = k\} \sum_{\ell=1}^{L_i} l_\tau \left[B_{\ell i} - \gamma^\top c\left(\frac{X_i - x}{h}\right) \right] K\left(\frac{x - X_i}{h}\right)$$

with respect to $\gamma \in \mathbf{R}^{|A_r|}$, where $l_\tau(u) \equiv \{|u| + (2\tau - 1)u\}/2$ for any $u \in \mathbf{R}$, and $K(\cdot)$ is a d-dimensional kernel function and h a bandwidth. More specifically, let $\widehat{q}_k(\tau | x) = \mathbf{e}_1^\top \hat{\gamma}_k(x)$, where $\hat{\gamma}_k(x) \equiv \arg\min_{\gamma \in \mathbf{R}^{|A_r|}} S_{n,x,\tau,k}(\gamma)$ and $\mathbf{e}_1$ is a column vector whose first entry is one, and the rest zero. Note that all bids are combined in each auction since we consider symmetric bidders.

6.4. Primitive Conditions

Let us present primitive conditions for the auction example of GPV. Let $\mathcal{P}$ denote the collection of the potential joint distributions of $(B^\top, X^\top, L)^\top$ and define $\mathcal{V} = \mathcal{T} \times \mathcal{P}$ as before.

For $u = (u_1, \ldots, u_d)^\top \in A_r$, and $r + 1$ times differentiable map f on $\mathbf{R}^d$, we define the following derivative:

$$(D^u f)(x) \equiv \frac{\partial^{[u]}}{\partial x_1^{u_1} \cdots \partial x_d^{u_d}} f(x),$$

where $[u] = u_1 + \cdots + u_d$. Then we define $\gamma_{\tau,k}(x) \equiv \left(\gamma_{\tau,k,u}(x)\right)_{u \in A_r}$, where

$$\gamma_{\tau,k,u}(x) \equiv \frac{1}{u_1! \cdots u_d!} D^u q_k(\tau|x).$$

In order to reduce the redundancy of the statements, let us introduce the following definitions.

DEFINITION 5. *Let $\mathcal{G}$ be a set of functions $g_v : \mathbf{R}^m \to \mathbf{R}^s$ indexed by a set $\mathcal{V}$, and let $S \subset \mathbf{R}^m$ be a given set and for $\varepsilon > 0$, let $S_v(\varepsilon)$ be an ε-enlargement of $S_v = \{x \in S : (x,v) \in S \times V\}$, i.e., $S_v(\varepsilon) = \{x + a : x \in S \text{ and } a \in [-\varepsilon, \varepsilon]^m\}$. Then we define the following conditions for $\mathcal{G}$:*

(a) *$B(S, \varepsilon)$: g_v is bounded on $S_v(\varepsilon)$ uniformly over $v \in \mathcal{V}$.*

(b) *$BZ(S, \varepsilon)$: g_v is bounded away from zero on $S_v(\varepsilon)$ uniformly over $v \in \mathcal{V}$.*

(c) *$BD(S, \varepsilon, r)$: $\mathcal{G}$ satisfies $B(S, \varepsilon)$ and g_v is r times continuously differentiable on $S_v(\varepsilon)$ with derivatives bounded on $S_v(\varepsilon)$ uniformly over $v \in \mathcal{V}$.*

(d) *$BZD(S, \varepsilon, r)$: $\mathcal{G}$ satisfies $BZ(S, \varepsilon)$ and g_v is r times continuously differentiable on $S_v(\varepsilon)$ with derivatives bounded on $S_v(\varepsilon)$ uniformly over $v \in \mathcal{V}$.*

(e) *LC: g_v is Lipschitz continuous with Lipschitz coefficient bounded uniformly over $v \in V$.*

Let $\mathcal{P}$ denote the collection of the potential joint distributions of $(B^\top, X^\top, L)^\top$ and define $\mathcal{V} = \mathcal{T} \times \mathcal{P}$, and for each $k \in \mathbb{N}_L$,

$$\mathcal{G}_q(k) = \{q_k(\tau|\cdot) : (\tau, P) \in \mathcal{V}\},$$
$$\mathcal{G}_f(k) = \{f_{\tau,k}(\cdot|\cdot) : (\tau, P) \in \mathcal{V}\},$$
$$\mathcal{G}_L(k) = \{P\{L_i = k|X_i = \cdot\} : P \in \mathcal{P}\}, \text{ and}$$
$$\mathcal{G}_f = \{f(\cdot) : P \in \mathcal{P}\}, \tag{6.4}$$

where $f_{\tau,k}(0|x)$ being the conditional density of $B_{li} - q_k(\tau|X_i)$ given $X_i = x$ and $L_i = k$. Also, define

$$\mathcal{G}_{f,2}(k) = \{f_{\cdot,k}(\cdot|\cdot) : P \in \mathcal{P}\} \text{ and } \mathcal{G}_\gamma(k) = \{\gamma_{\cdot,k}(\cdot) : P \in \mathcal{P}\}. \tag{6.5}$$

We make the following assumptions.

Assumption AUC1.   (i) $\mathcal{G}_f$ satisfies BD($\mathcal{S}, \varepsilon, 1$).

(ii) For each $k \in \mathbb{N}_L$, $\mathcal{G}_f(k)$ and $\mathcal{G}_L(k)$ satisfy BD($\mathcal{S}, \varepsilon, 1$) and BZD($\mathcal{S}, \varepsilon, 1$).

(iii) For each $k \in \mathbb{N}_L$, $\mathcal{G}_q(k)$ satisfies BD($\mathcal{S}, \varepsilon, r+1$) for some $r > 3d/2 - 1$.

(iv) For each $k \in \mathbb{N}_L$, $\mathcal{G}_{f,2}(k)$ and $\mathcal{G}_\gamma(k)$ satisfy LC.

Assumption AUC1(i) and (iii) are standard assumptions used in the local polynomial approach where one approximates $q_k(\cdot|x)$ by a linear combination of its derivatives through Taylor expansion, except only that the approximation here is required to behave well uniformly over $P \in \mathcal{P}$. Assumption AUC1(ii) is made to

prevent the degeneracy of the asymptotic linear representation of $\hat{\gamma}_{\tau,k}(x) - \gamma_{\tau,k}(x)$ that is uniform over $x \in \mathcal{S}_\tau(\varepsilon)$, $\tau \in \mathcal{T}$ and over $P \in \mathcal{P}$. Assumption AUC(iv) requires that the conditional density function of $B_{li} - q_k(\tau|X_i)$ given $X_i = x$ and $L_i = k$ and $\gamma_{\tau,k}(\cdot)$ behave smoothly as we perturb τ locally. This requirement is used to control the size of the function spaces indexed by τ, so that when the stochastic convergence of random sequences holds, it is ensured to hold uniformly in τ.

Assumption AUR2 lists conditions for the kernel function and the bandwidth.

> **Assumption AUC2.**   (i) K is compact-supported, nonnegative, bounded, and Lipschitz continuous on the interior of its support, $\int K(u)du = 1$, and $\int K(u)\|u\|^2 du > 0$.
>
> (ii) $n^{-1/2}h^{-3(d+v)/2} + \sqrt{n}h^{r+d+1}/\sqrt{\log n} \to 0$, as $n \to \infty$, for some small $v > 0$, with r in Assumption AUC1(iii).

As for Assumption AUC2(ii), the choice of $h = n^{-s}$ with the condition $1/(2(r+d+1)) < s < 1/(3(d+v))$ satisfies the bandwidth condition. The small $v > 0$ there is introduced to satisfy Assumption B4.

> **Assumption AUC3.** $\hat{\underline{b}} = \underline{b} + o_P\left(n^{-1/2}\right)$, $\mathcal{P}$-uniformly.

> **Assumption AUC4.**   (i) There exist nonstochastic sequences $c_{n,L} > 0$ and $c_{n,U} > 0$ such that $c_{n,L} < c_{n,U}$, and as $n \to \infty$,
>
> $$\inf_{P \in \mathcal{P}} P\left\{\hat{c}_n \in [c_{n,L}, c_{n,U}]\right\} \to 1, \text{ and } \sqrt{\log n}/c_{n,L} + n^{-1/2}h^{-d/2}c_{n,U} \to 0.$$

(ii) For each $a \in (0, 1/2)$, there exists a compact set $\mathcal{C}_a \subset \mathbf{R}^d$ such that

$$0 < \inf_{P \in \mathcal{P}} P\{X_i \in \mathbf{R}^d \backslash \mathcal{C}_a\} \le \sup_{P \in \mathcal{P}} P\{X_i \in \mathbf{R}^d \backslash \mathcal{C}_a\} < a.$$

Assumption AUC3 holds in general because the extreme order statistic is super-consistent with the n^{-1} rate of convergence. Assumption AUC4(i) requires that $\hat{c}_n$ increase faster than $\sqrt{\log n}$ but slower than r_n with probability approaching one. Assumption AUC4(ii) imposes some regularity on the behavior of the support of X_i as P moves around $\mathcal{P}$.

The following result establishes the uniform validity of the bootstrap test.

THEOREM AUC1. *Suppose that Assumptions AUC1–AUC4 hold. Then*

$$\limsup_{n \to \infty} \sup_{P \in \mathcal{P}_0} P\left\{\hat{\theta}_{sum} > c^*_{\alpha,\eta}\right\} \le \alpha \text{ and } \limsup_{n \to \infty} \sup_{P \in \mathcal{P}_0} P\left\{\hat{\theta}_{max} > c^*_{\alpha,\eta}\right\} \le \alpha.$$

Theorem AUC1 gives the uniform asymptotic validity of the bootstrap test. It is straightforward to characterize the class of distributions under the null hypothesis which renders the test asymptotically exact, using Theorem 2. We omit the details for brevity.

Our asymptotic approximation is based on plugging the asymptotic linear expansion directly. There is a recent proposal by Mammen, Van Keilegom, and Yu (2013), who developed nonparametric tests for parametric specifications of regression quantiles and showed that calculating moments of linear expansions of nonparametric quantile regression estimators might work better in a sense that their approach requires less stringent conditions for the dimension of covariates and the choice of the bandwidth. It is an interesting future research topic whether their ideas can be applied to our setup.

6.5. Empirical Results

We now present empirical results using the timber auction data used in Lu and Perrigne (2008).[20] They used the timber auction data to estimate bidders' risk aversion, taking advantage of bidding data from ascending auctions as well as those from first-price sealed-bid auctions. In our empirical example, we use only the latter auctions with 2 and 3 bidders, and we use the appraisal value as the only covariate X_i $(d = 1)$. Summary statistics and visual presentation of data are given in Table 3 and Figure 2. It can be seen from Table 3 that average bids become higher as the number of bidders increases from 2 to 3. The top panel of Figure 2 suggests that this more aggressive bidding seems to be true, conditional on appraisal values.

Before estimation, the covariate was transformed to lie between 0 and 1 by studentizing it and then applying the standard normal CDF transformation. The bottom panel of Figure 2 shows local linear estimates of conditional quantile functions at $\tau = 0.1, 0.5, 0.9$.[21] In this figure, estimates are only shown between the 10% and 90% sample quantiles of the covariate.

On one hand, the 10% conditional quantiles are almost identical between auctions with two bidders $(I = 2)$ and those with three bidders $(I = 3)$. On the other hand, the 50% and 90% conditional quantiles are higher with three bidders for most values of appraisal values. There is a crossing of two conditional median curves at the lower end of appraisal values.

TABLE 3. Summary statistics for empirical example 1

	2 Bidders (sample size = 107)		3 Bidders (sample size = 108)	
	Mean	Standard deviation	Mean	Standard deviation
Appraisal value	66.0	47.7	53.3	41.4
Highest bid	96.1	55.6	100.8	56.7
Second highest bid	80.9	49.2	83.1	51.5
Third highest bid			69.4	44.6

Notes: Bids and appraisal values are given in dollars per thousand board-feet (MBF).
Source: Timber auction data are from the *Journal of Applied Econometrics* website.

Data Points

Conditional Quantiles of All Bids
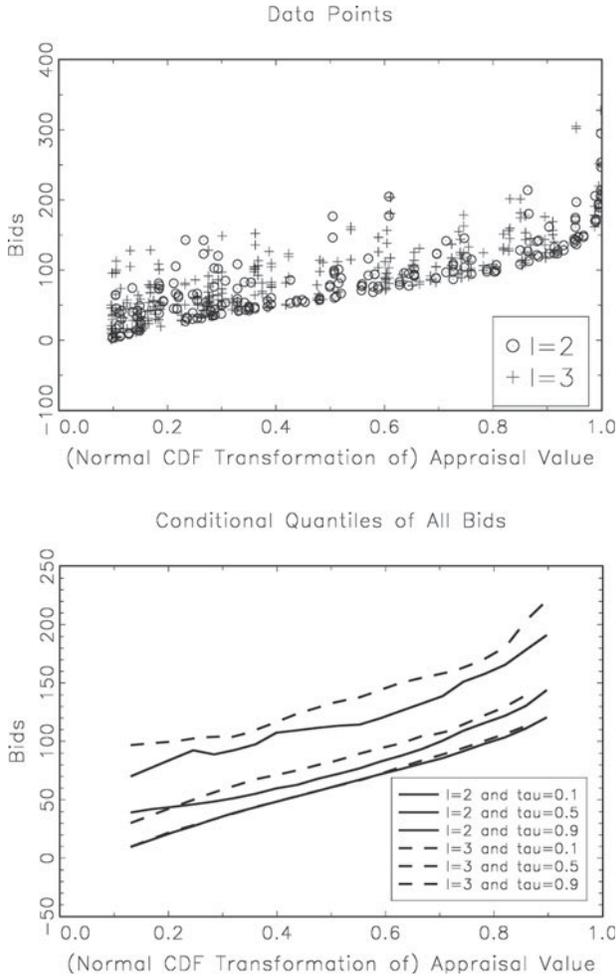
FIGURE 2. Data for empirical illustration for empirical example 1.

Note: The top panel of the figure shows observations and the bottom panel depicts local linear quantile regression estimates.

To check whether inequalities in (6.2) hold in this empirical example, we plot estimates of $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$ in Figure 3. The top panel of the figure shows that 20 estimated curves of $v_{\tau,1}(x)$, each representing a particular conditional quantile, ranging from the 10th percentile to the 90th percentile. There are strictly positive values of $v_{\tau,1}(x)$ at the lower end of appraisal values. The bottom panel of Figure 3 depicts 20 estimated curves of $v_{\tau,2}(x)$, showing that they are all strictly negative. The test based on (6.3) can tell formally whether positive values of $v_{\tau,1}(x)$ at the lower end of appraisal values can be viewed as evidence against economic restrictions imposed by (6.2).

q_1(taulx) − q_2(taulx)

(Normal CDF Transformation of) Appraisal Value

bid_lower_bound − 2*q_1(taulx) + q_2(taulx)

(Normal CDF Transformation of) Appraisal Value
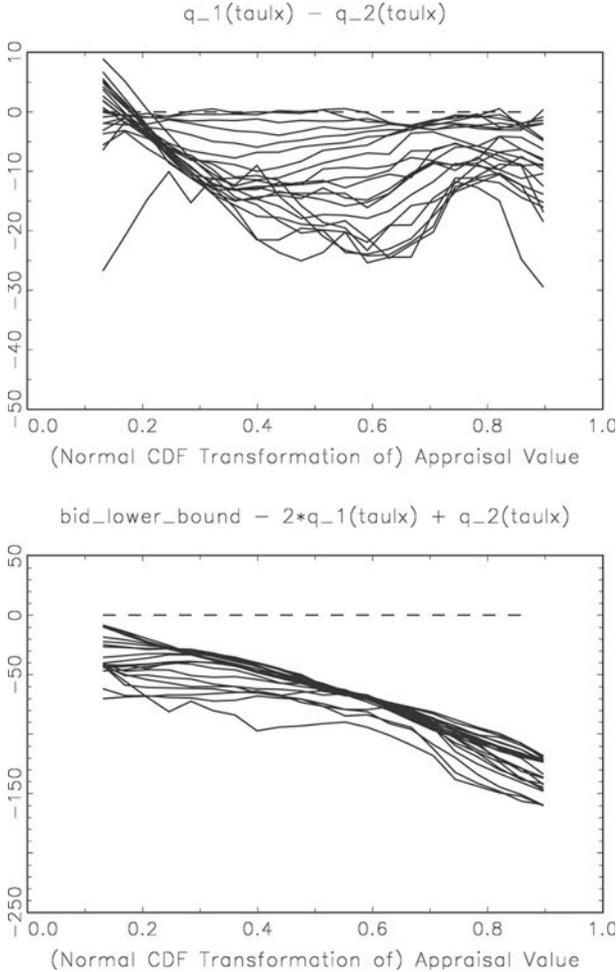
FIGURE 3. Estimates of $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$ for empirical example 1.

Note: The top and bottom panels of the figure show estimates of $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$, respectively, where $\hat{v}_{\tau,1}(x) = \hat{q}_1(\tau|x) - \hat{q}_2(\tau|x)$ and $\hat{v}_{\tau,2}(x) = \underline{b} - 2\hat{q}_1(\tau|x) + \hat{q}_2(\tau|x)$.

We considered both the L_1 and L_2 test statistics described in (6.3). We set $\mathcal{T}$ to be the interval between the 10th and 90th percentiles of the covariate, and also set $\mathcal{X} = [0.1, 0.9]$. The contact set was estimated with $\hat{c}_n = C_{cs} \log\log(n) q_{1-0.1/\log(n)}(S_n^*)$ with $r_n = \sqrt{nh}$. We checked the sensitivity to the tuning parameters with $C_{cs} \in \{0.5, 1, 1.5\}$ and $h \in \{0.3, 0.6, 0.9\}$. All cases resulted in bootstrap p-values of 1, thereby suggesting that positive values of $v_{\tau,1}(x)$ at the lower end of appraisal values cannot be interpreted as evidence against the null hypothesis beyond random sampling errors. Therefore, we have not found any evidence against economic implications imposed by (6.2).

7. EMPIRICAL EXAMPLE 2: TESTING FUNCTIONAL INEQUALITIES IN THE CONTEXT OF WAGE INEQUALITY

In this section, we give an empirical example regarding testing functional inequalities via differences-in-differences in conditional quantiles, inspired by Acemoglu and Autor (2011).

7.1. Testing Problem

Figures 9a–9c in Acemoglu and Autor (2011) depict changes in log hourly wages by percentile relative to the median. Specifically, they consider the following differences-in-differences in quantiles:

$$\Delta_{t,s}(\tau, x) \equiv [q_t(\tau|x) - q_s(\tau|x)] - [q_t(0.5|x) - q_s(0.5|x)]$$

for time periods t and s and for quantiles τ, where $q_t(\tau|x)$ denotes the τ-quantile of log hourly wages conditional on $X = x$ in year t. Acemoglu and Autor (2011) consider males and females together in Figure 9a, males only in Figure 9b, and females only in Figure 9c. Thus, in their setup, the only covariate X is gender.

Figures 9a–9c in Acemoglu and Autor (2011) suggest that (1) $\Delta_{1988,1974}(\tau, x) \geq 0$ for quantiles above the median, but $\Delta_{1988,1974}(\tau, x) \leq 0$ for quantiles below the median (hence, widening the wage inequality, while the lower quantiles are losing most), and that (2) $\Delta_{2008,1988}(\tau, x) \geq 0$ for most quantiles (hence, 'polarization' of wage growth, while the middle quantiles lose most). In this subsection, we consider testing

$$H_0 : \Delta_{t,s}(\tau, x) \geq 0 \; \forall (x, \tau) \in \mathcal{X} \times \mathcal{T}, \tag{7.1}$$

with a continuous covariate (age in our empirical example), where $(t, s) = (1988, 1974)$ or $(t, s) = (2008, 1988)$.[22] Note that degeneracy of the test statistic could occur if the contact set consists of values of (x, τ) only around $\tau = 0.5$. Therefore, the uniformity of our test could be potentially important in this example.

7.2. Test Statistic

To implement the test, we again use a local polynomial quantile regression estimator, say $\hat{q}_t(\tau|x)$. Then $\Delta_{t,s}(\tau, x)$ can be estimated by

$$\hat{\Delta}_{t,s}(\tau, x) \equiv [\hat{q}_t(\tau|x) - \hat{q}_s(\tau|x)] - [\hat{q}_t(0.5|x) - \hat{q}_s(0.5|x)].$$

Then testing (7.1) can be carried out using

$$\hat{\theta}_{t,s} \equiv \int_{\mathcal{X} \times \mathcal{T}} [r_n \hat{v}_{\tau,t,s}(x)]_+^p \, dQ(x, \tau), \tag{7.2}$$

where $\hat{v}_{\tau,t,s}(x) = -\hat{\Delta}_{t,s}(\tau, x)$.[23] Here, to reflect different sample sizes between two time periods, we set

$$r_n = \sqrt{\frac{(n_t h_t) \times (n_s h_s)}{(n_t h_t) + (n_s h_s)}},$$

where n_j and h_j are the sample size and the bandwidth used for nonparametric estimation for year $j = t, s$.

7.3. Empirical Results

We used the CPS data extract of Acemoglu and Autor (2011).[24] In our empirical example, we use age in years as the only covariate. Summary statistics and visual presentation of data are given in Table 4 and Figure 4. Note that Figure 4 replicates the basic patterns of Figures 9 of Acemoglu and Autor (2011).

TABLE 4. Summary statistics for empirical example 2

Year	1974	1988	2008
Log real hourly wages	2.780	2.769	2.907
Age in years	35.918	35.501	39.051
Sample size	19,575	64,682	48,341

Notes: The sample is restricted to white males aged between 16 and 64. Entries for log real hourly wages and age show CPS sample weighted means. *Source:* May/ORG CPS data extract from David Autor's web site.
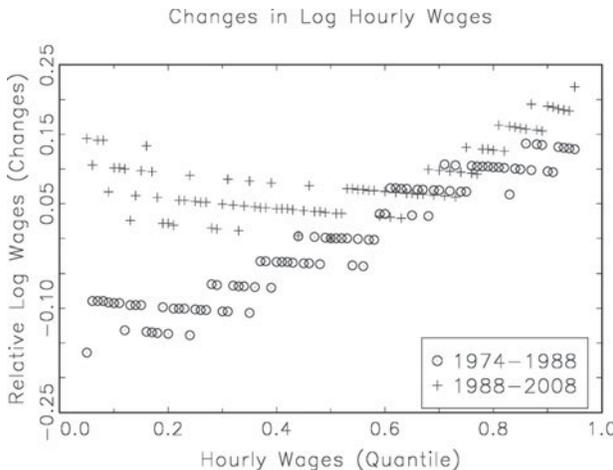
FIGURE 4. Changes in log hourly wages by percentile relative to the median.

Notes: The figure shows differences-in-differences in quantiles of log hourly wages, measured by $[q_t(\tau) - q_s(\tau)] - [q_t(0.5) - q_s(0.5)]$. Triangles correspond to changes from 1974 to 1988, whereas circles signify those from 1988 to 2008. All quantiles are computed using CPS sample weight. *Source:* May/ORG CPS data extract from David Autor's web site.

We now turn to the conditional version of Figure 4, using age as a conditioning variable. As an illustration, let $\mathcal{X}$ be an interval of ages between 25 and 60 and let $\mathcal{T} = [0.1, 0.9]$. To check whether inequalities in $\hat{\Delta}_{t,s}(\tau, x) \geq 0$ hold for each value of $(x, \tau) \in \mathcal{X} \times \mathcal{T}$, we plot estimates of $\hat{\upsilon}_{\tau,t,s}(x) = -\hat{\Delta}_{t,s}(\tau, x)$ in Figure 5. The top panel of the figure shows that 5 estimated curves of $\hat{\upsilon}_{\tau,1988,1974}(x)$, each representing a particular conditional quantile, and the bottom panel shows the corresponding graph for period 1988–2008.[25] By construction, the estimated curve is a flat line at zero when $\tau = 0.5$. As consistent with Figure 4, the lower quantiles
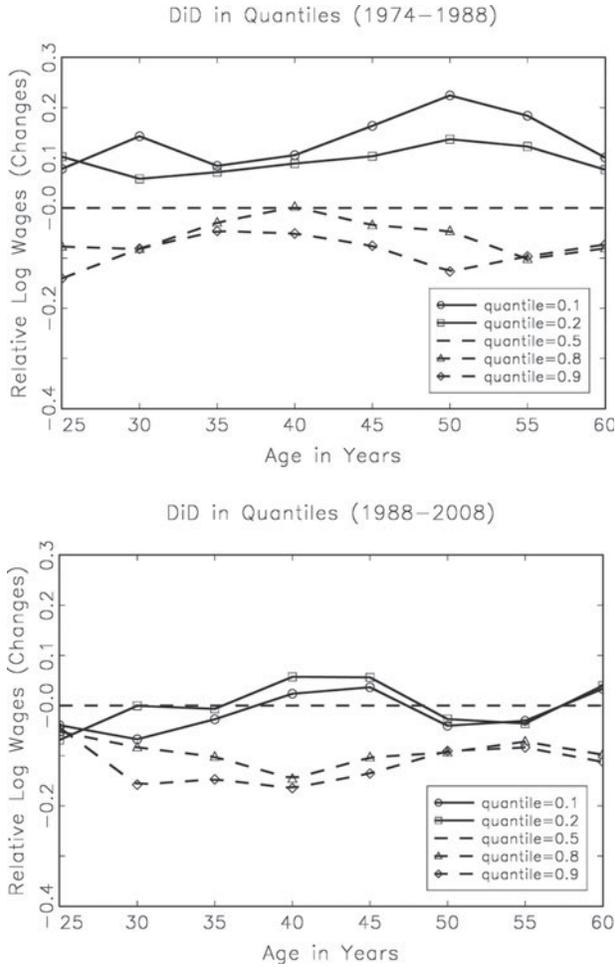
FIGURE 5. Estimates of $\hat{\upsilon}_{\tau,t,s}(x)$.

Note: The top and bottom panels of the figure show local linear estimates of $-\Delta_{1988,1974}(\tau, x)$ and $-\Delta_{2008,1988}(\tau, x)$, respectively, where x is age in years.

seem to violate the null hypothesis, especially for the period 1974–1988. As before, our test can tell formally whether positive values of $\hat{v}_{\tau,t,s}(x)$ lead to rejection of the null hypothesis of polarization of wage growth.

We considered both the L_1 and L_2 test statistics described in (7.2). As before, the contact set was estimated with $\hat{c}_n = C_{\mathrm{cs}} \log\log(n) q_{1-0.1/\log(n)}(S_n^*)$ with $r_n = \sqrt{nh}$.[26] We checked the sensitivity to the tuning parameters with $C_{\mathrm{cs}} \in \{0.5, 1, 1.5\}$.

For period 1974–1988, we rejected the null hypothesis at the 1% level across all three values of C_{cs}. However, for period 1988–2008, we fail to reject the null hypothesis at the 5% level for any value of C_{cs}. Therefore, the changing patterns of the US wage distribution around 1988, reported in Acemoglu and Autor (2011), seem to hold up conditionally on age as well.

8. CONCLUSIONS

In this paper, we have proposed a general method for testing inequality restrictions on nonparametric functions and have illustrated its usefulness by looking at two particular empirical applications. We regard our examples as just some illustrative applications and believe that our framework can be useful in a number of other settings.

Our bootstrap test is based on a one-sided version of L_p functionals of kernel-type estimators ($1 \le p < \infty$). We have provided regularity conditions under which the bootstrap test is asymptotically valid uniformly over a large class of distributions and have also provided a class of distributions for which the asymptotic size is exact. We have shown the consistency of our test and have obtained a general form of the local power function.

There are different notions of efficiency for nonparametric tests and hence there is no compelling sense of an asymptotically optimal test for the hypothesis considered in this paper. See Nikitin (1995) and Bickel, Ritov, and Stoker (2006) for general discussion. However, if we restrict our attention to inference on a finite dimensional parameter in a model defined by conditional moment inequalities, it follows from the recent results Armstrong (2014a,b) in the literature that the multiscale tests of Armstrong and Chan (2016) and Chetverikov (2017) achieve adaptive rate-optimality against a sequence of smooth alternatives and the test in Chernozhukov et al. (2013) is also rate-optimal under appropriate bandwidth choices for the bandwidth.

We have used the fixed constant η to establish the uniformity, while allowing for degenerate cases. It is a topic for future research to investigate whether one can get rid of this tuning parameter by using a different scaling or some alternative methods.

NOTES

1. Permitting the convergence rate $r_{n,j}$ to differ across $j \in \mathbb{N}_J$ can be convenient, when the nonparametric estimators have different convergence rates. For example, this accommodates a situation where one jointly tests the nonnegativity and monotonicity of a nonparametric function.

2. While our framework permits the case where $\hat{\sigma}_{\tau,j}(x)$ is simply chosen to be 1, we allow for a more general case where $\hat{\sigma}_{\tau,j}(x)$ is a consistent estimator for some nonparametric quantity.

3. It is convenient for general development to let the population quantities $v_{n,\tau,j}(x)$ and $\sigma_{n,\tau,j}(x)$ depend on n.

4. We fix the value of η for the precise definition of the test statistic; however, its value does not matter in terms of the first-order asymptotic theory.

5. For example, take $\sqrt{nh}v_{n,1}(x) = -x^2/n$ on $\mathcal{X} = [-1, 1]$. Let $v_0(x) \equiv 0$. Then $\sqrt{nh}v_{n,1}(x)$ goes to $v_0(x)$ uniformly in $x \in \mathcal{X}$ as $n \to \infty$. However, for each n, $B_{n,1}(0) = \{x \in \mathcal{X} : \sqrt{nh}v_{n,1}(x) = 0\} = \{0\}$, which does not converge in Hausdorff distance to $B_1(0) \equiv \{x \in \mathcal{X} : v_0(x) = 0\} = \mathcal{X}$.

6. Our fixed positive constant η plays a role similar to a fixed constant in Andrews and Shi (2013)'s modification of the sample variance-covariance matrix of unconditional moment conditions, transformed by instruments (ε in their notation in equation (3.5) of Andrews and Shi (2013)).

7. The uniformity result in this paper is nonstandard since our test is based on asymptotically non-tight processes, in contrast to Andrews and Shi (2013) who convert conditional moment inequalities into an infinite number of unconditional moment inequalities. This paper's development of asymptotic theory draws on the method of Poissonization (see, e.g., Rosenblatt (1975), Horváth (1991) and Giné, Mason, and Zaitsev (2003)). For applications of this method, see Anderson, Linton, and Whang (2012) for inference on a polarization measure, Chang et al. (2015) for testing for conditional treatment effects, and Lee et al. (2013) for testing inequalities for nonparametric regression functions using the numerator of the Nadaraya-Watson estimator (based on pointwise asymptotics). Also, see Mason and Polonik (2009) and Biau, Cadre, Mason, and Pelletier (2009) for support estimation.

8. In fact, the main challenge here is to prove the bootstrap approximation using the method of Poissonization that is uniform in $P \in \mathcal{P}_0$.

9. Throughout the paper, we assume that $X_i \in \mathbf{R}^d$ is a continuous random vector. It is straightforward to extend the analysis to the case where X_i has a subvector of discrete random variables.

10. To see this more clearly, we assume that $\mathcal{T} = \{\tau\}$, $p = 1$, and $J = 1$, and suppress the subscripts τ and j from the notation, and take $\hat{\sigma}(x) = 1$ for simplicity. We write (in the case where $v_n(x) = 0$)

$$h^{-d/2}\hat{\theta} = h^{-d/2}\int_{\mathcal{X}} \max\left\{r_n\{\hat{v}(x) - v_n(x)\}, 0\right\} dx$$

$$= h^{-d/2}\int_{\mathcal{X}} \max\left\{\sqrt{nh^d}\{\hat{g}(x) - \mathbf{E}\hat{g}(x)\}, 0\right\} dx + h^{-d/2}R_n,$$

where R_n is an error term that has at least the same convergence rate as the convergence rate of the remainder term in the asymptotic linear representation for $\hat{v}(x)$. Now we let

$$a_n = \mathbf{E}\left[\int_{\mathcal{X}} \max\left\{\sqrt{nh^d}\{\hat{g}(x) - \mathbf{E}\hat{g}(x)\}, 0\right\} dx\right]$$

and write $h^{-d/2}\hat{\theta} - h^{-d/2}a_n$ as

$$h^{-d/2}\left(\int_{\mathcal{X}} \max\left\{\sqrt{nh^d}\{\hat{g}(x) - \mathbf{E}\hat{g}(x)\}, 0\right\} dx - a_n\right) + h^{-d/2}R_n.$$

It can be shown that the leading term is asymptotically normal using the method of Poissonization. Hence $h^{-d/2}\hat{\theta} - h^{-d/2}a_n$ becomes asymptotically normal, if $R_n = o_P(h^{d/2})$. This is where the faster error rate in the asymptotic linear representation in Assumption A1(i) plays a role.

11. The conditional expectation $\mathbf{E}_P\left[|\beta_{n,x,\tau,j}(Y_{ij}, u)|^M | X_i = x\right]$ is of type $\mathbf{E}[f(Y, x)|X = x]$, which is not well defined according to Kolmogorov's definition of conditional expectations. See, e.g., Proschan and Presnell (1998) for this problem. Here we define the conditional expectation in an elementary way by using conditional densities or conditional probability mass functions of (Y_{ij}, Y_{ik}) given $X_i = x$, depending on whether (Y_{ij}, Y_{ik}) is continuous or discrete.

12. (Chen, Linton, and Van Keilegom, 2003, Theorem 3) introduced its extension to functions indexed partly by infinite dimensional parameters, and called it local uniform L_2-continuity. For further discussions, see Andrews (1994) and Chen et al. (2003).

13. As general discussions, our test involves a one-sided version of L_p-type functionals of nonparametric estimators ($1 \leq p < \infty$). We regard the sup-norm and L_p norm approaches as complementary, each with its own strength and weakness. For example, our test and also the test of Andrews and Shi (2013) have higher power against relatively flat alternatives, whereas the test of Chernozhukov et al. (2013) has higher power against sharply-peaked alternatives. See the results of Monte Carlo experiments reported in Appendix 5. See also Andrews and Shi (2013), Andrews and Shi (2014), and Chernozhukov et al. (2013) for related discussions and further Monte Carlo evidence. Rosenblatt (1975) provides the comparison between alternative local sequences in the context of a nonparametric test of goodness of fit for the density function.

14. The local power results in this section are more general than those of Lee et al. (2013). In particular, the results accommodate a wider class of local alternatives that may not converge to the least favorable case.

15. Note that $q_n(S_n^*)$ is the $(1 - 0.1/\log n)$ quantile of the supremum of $\hat{s}_{\tau,j}^*(x)$ over (j, τ, x) and that $(1 - 0.1/\log n)$ converges to 1 as n gets large. Thus, this observation leads to the choice of $\hat{c}_n$ in (5.1) that is proportional to $q_n(S_n^*)$ times a very slowing growing term such as $\log\log n$, to insure that $\hat{c}_n$ diverges to infinity but as slowly as possible, while having the property of scale invariance.

16. See Assumption A4(ii) below for sufficient conditions for a data dependent choice of $\hat{c}_n$. It is not hard to see that the conditions are satisfied, once the uniform convergence rates of $\hat{v}_{\tau,j}(x)$ and $\hat{\sigma}_{\tau,j}(x)$ and their bootstrap versions hold as required in Assumptions A3, A5, and B2 and B3.

17. In GPV, it is assumed that for $I = k$, the support of the observed equilibrium bid distribution is $[\underline{b}, \overline{b}_k] \subset [0, \infty)$ with $\underline{b} < \overline{b}_k$, where $k = 2, 3$. Note that $\underline{b}$ is common across k's, while $\overline{b}_k$'s are not.

18. If necessary, we may test the strict inequalities (3.1), instead of the weak inequalities (3.2). However, such a test would require a test statistic that is different from ours and needs a separate treatment.

19. In our application, we set $\hat{\underline{b}}$ to be the observed minimum value.

20. The data are available on the *Journal of Applied Econometrics* website.

21. Specifically, the conditional quantile functions $q_2(\tau|x)$ and $q_3(\tau|x)$ are estimated via the local linear quantile regression estimator with the kernel function $K(u) = 1.5[1 - (2u)^2] \times 1\{|u| \leq 0.5\}$ and the bandwidth $h = 0.6$. See Section 6.3 for more details on estimating conditional quantile functions.

22. Note that H_0 in (7.1) includes the case $\Delta_{t,s}(\tau, x) \equiv 0$, which does not correspond to the notion of polarization. In view of this, our null hypothesis in (7.1) can be regarded as a weak form of polarization hypothesis, whereas a more strict version can be written as the inequality in (7.1) holds strictly for some high and low quantiles.

23. Note that the null hypothesis is written as positivity in (7.1). Hence $\hat{v}_{\tau,t,s}(x)$ is defined accordingly.

24. The data are available on David Autor's web site. We would like to thank him for posting the data set on a public domain. They used three-year averages around the year of interest to produce Figures 9a–9c in Acemoglu and Autor (2011); however, we used just annual data.

25. As before, underlying conditional quantile functions are estimated via the local linear quantile regression estimator with the kernel function $K(u) = 1.5[1 - (2u)^2] \times 1\{|u| \leq 0.5\}$. One important difference from the first empirical example is that we used the CPS sample weight, which were incorporated by multiplying it to the kernel weight for each observation. Finally, the bandwidth was $h = 2.5$ for all years.

26. To accommodate different sample sizes across years, we set $n = (n_{1974} + n_{1988} + n_{2008})/3$ in computing $\hat{c}_n$. The bandwidth was $h = 2.5$.

REFERENCES

Acemoglu, D. & D. Autor (2011) Skills, tasks and technologies: Implications for employment and earnings. In D. Card & O. Ashenfelter (eds.), *Handbook of Labor Economics*, vol. 4, part B, pp. 1043–1171. Elsevier.

Anderson, G., O. Linton, & Y.-J. Whang (2012) Nonparametric estimation and inference about the overlap of two distributions. *Journal of Econometrics* 171(1), 1–23.

Andrews, D.W.K. (1994) Empirical process method in econometrics. In R.F. Engle & D.L. McFadden (eds.), *The Handbook of Econometrics*, vol. 4, pp. 2247–2294. North-Holland.

Andrews, D.W.K. & X. Shi (2013) Inference based on conditional moment inequalities. *Econometrica* 81(2), 609–666.

Andrews, D.W.K. & X. Shi (2014) Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics* 179(1), 31–45.

Andrews, D.W.K. & X. Shi (2017) Inference based on many conditional moment inequalities. *Journal of Econometrics* 196(2), 275–287.

Aradillas-López, A., A. Gandhi, & D. Quint (2013) Identification and inference in ascending auctions with correlated private values. *Econometrica* 81(2), 489–534.

Aradillas-López, A., A. Gandhi, & D. Quint (2016) A simple test for moment inequality models with an application to English auctions. *Journal of Econometrics* 194(1), 96–115.

Armstrong, T.B. (2014a) On the choice of test statistic for conditional moment inequalities. ArXiv:1410.4718.

Armstrong, T.B. (2014b) Weighted KS statistics for inference on conditional moment inequalities. *Journal of Econometrics* 181(2), 92–116.

Armstrong, T.B. (2015) Asymptotically exact inference in conditional moment inequalitymodels. *Journal of Econometrics* 186(1), 51–65.

Armstrong, T.B. & H.P. Chan (2016) Multiscale adaptive inference on conditional moment inequalities. *Journal of Econometrics* 194(1), 24–43.

Baraud, Y., S. Huet, & B. Laurent (2005) Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Annals of Statistics* 33(1), 214–257.

Beare, B.K. & J.-M. Moon (2015) Nonparametric tests of density ratio ordering. *Econometric Theory* 31(3), 471–492.

Biau, G., B. Cadre, D. Mason, & B. Pelletier (2009) Asymptotic normality in density support estimation. *Electronic Journal of Probability* 14(91), 2617–2635.

Bickel, P.J., Y. Ritov, & T.M. Stoker (2006) Tailor-made tests for goodness of fit to semiparametric hypotheses. *Annals of Statistics* 34(2), 721–741.

Chang, M., S. Lee, & Y.-J. Whang (2015) Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements. *Econometrics Journal* 18(3), 307–346.

Chen, X., O. Linton, & I. Van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71(5), 1591–1608.

Chernozhukov, V., S. Lee, & A. Rosen (2013) Intersection bounds: Estimation and inference. *Econometrica* 81(2), 667–737.

Chetverikov, D. (2012) Testing Regression Monotonicity in Econometric Models. Working paper, MIT.

Chetverikov, D. (2017) Adaptive tests of conditional moment inequalities. *Econometric Theory*, forthcoming. doi:10.1017/S0266466617000184.

Delgado, M.A. & J.C. Escanciano (2012) Distribution-free tests of stochastic monotonicity. *Journal of Econometrics* 170(1), 68–75.

Delgado, M.A. & J.C. Escanciano (2013) Conditional stochastic dominance testing. *Journal of Business & Economic Statistics* 31(1), 16–28.

Dümbgen, L. & V.G. Spokoiny (2001) Multiscale testing of qualitative hypotheses. *Annals of Statistics* 29(1), 124–152.

Fan, Y. & S. Park (2014) Nonparametric inference for counterfactual means: Bias-correction, confidence sets, and weak IV. *Journal of Econometrics* 178(Part 1), 45–56.

Ghosal, S., A. Sen, & A.W. van der Vaart (2000) Testing monotonicity of regression. *Annals of Statistics* 28, 1054–1082.

Gimenes, N. & E. Guerre (2013) Augmented Quantile Regression Methods for First-Price Auction. Working paper, Queen Mary University of London.

Giné, E., D.M. Mason, & A.Y. Zaitsev (2003) The L_1-norm density estimator process. *Annals of Probability* 31, 719–768.

Guerre, E., I. Perrigne, & Q. Vuong (2009) Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica* 77(4), 1193–1227.

Haile, P., H. Hong, & M. Shum (2003) Nonparametric Tests for Common Values at First- Price Sealed-Bid Auctions. Working paper, Yale University.

Haile, P.A. & E. Tamer (2003) Inference with an incomplete model of English auctions. *Journal of Political Economy* 111(1), 1–51.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21(4), 1926–1947.

Horváth, L. (1991) On L_p-norms of multivariate density estimators. *Annals of Statistics* 19(4), 1933–1949.

Hsu, Y.-C. (2017) Consistent tests for conditional treatment effects. *Econometrics Journal* 20, 1–22.

Juditsky, A. & A. Nemirovski (2002) On nonparametric tests of positivity/monotonicity/convexity. *Annals of Statistics* 30(2), 498–527.

Jun, S.J., J. Pinkse, & Y. Wan (2010) A consistent nonparametric test of affiliation in auction models. *Journal of Econometrics* 159(1), 46–54.

Khan, S. & E. Tamer (2009) Inference on endogenously censored regression models using conditional moment inequalities. *Journal of Econometrics* 152(2), 104–119.

Krasnokutskaya, E., K. Song, & X. Tang (2016) The Role of Quality in On-Line Service Markets. Working paper, Johns Hopkins University.

Lee, S., O. Linton, & Y.-J. Whang (2009) Testing for stochastic monotonicity. *Econometrica* 77(2), 585–602.

Lee, S., K. Song, & Y.-J. Whang (2013) Testing functional inequalities. *Journal of Econometrics* 172(1), 14–32.

Linton, O., K. Song, & Y.-J. Whang (2010) An improved bootstrap test of stochastic dominance. *Journal of Econometrics* 154(2), 186–202.

Lu, J. & I. Perrigne (2008) Estimating risk aversion from ascending and sealed-bid auctions: The case of timber auction data. *Journal of Applied Econometrics* 23(7), 871–896.

Mammen, E., I. Van Keilegom, & K. Yu (2013) Expansion for moments of regression quantiles with application to nonparametric testing. Working paper, arXiv:1306.6179.

Marmer, V., A. Shneyerov, & P. Xu (2013) What model for entry in first-price auctions? A nonparametric approach. *Journal of Econometrics* 176(1), 46–58.

Mason, D.M. & W. Polonik (2009) Asymptotic normality of plug-in level set estimates. *Annals of Applied Probability* 19, 1108–1142.

Menzel, K. (2014) Consistent estimation with many moment inequalities. *Journal of Econometrics* 182(2), 329–350.

Nikitin, Y. (1995) *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press.

Proschan, M.A. & B. Presnell (1998) Expect the unexpected from conditional expectation. *American Statistician* 52(3), 248–252.

Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics* 3(1), 1–14.