Refinements of the Kiefer-Wolfowitz theorem and a test of concavity[*]

Zheng Fang

Department of Economics
Texas A&M University
College Station, TX 77843
e-mail: zfang@tamu.edu

Abstract: This paper studies estimation of and inference on a distribution function F that is concave on the nonnegative half line and admits a density function f with potentially unbounded support. When F is strictly concave, we show that the supremum distance between the Grenander distribution estimator and the empirical distribution may still be of order $O(n^{-2/3}(\log n)^{2/3})$ almost surely, which reduces to an existing result of Kiefer and Wolfowitz when f has bounded support. We further refine this result by allowing F to be not strictly concave or even non-concave and instead requiring it be "asymptotically" strictly concave. Building on these results, we then develop a test of concavity of F or equivalently monotonicity of f, which is shown to have asymptotically pointwise level control under the entire null as well as consistency under any fixed alternative. In fact, we show that our test has local size control and nontrivial local power against any local alternatives that do not approach the null too fast, which may be of interest given the irregularity of the problem. Extensions to settings involving testing concavity/convexity/monotonicity are discussed.

MSC 2010 subject classifications: 62G05, 62G07, 62G09, 62G10, 62G20.
Keywords and phrases: Grenander estimator, Kiefer-Wolfowitz theorem, least concave majorant, test of concavity/monotonicity.

Received August 2018.

1. Introduction

Let $\{X_i\}_{i=1}^n$ be a sample of i.i.d. nonnegative random variables with common distribution function F that is known to be concave. The seminal work by Grenander (1956) establish that the nonparametric maximum likelihood estimator of F is given by the least concave majorant $\hat{\mathbb{F}}_n$ of the empirical distribution function $\mathbb{F}_n$. In this context, the present paper provides new results on two aspects of the problem.

The first set of results concerns the closeness between $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ relative to the uniform norm $\|\cdot\|_\infty$. The celebrated Marshall's lemma indicates that $\hat{\mathbb{F}}_n$ is

---

$\sqrt{n}$-consistent for F relative to $\| \cdot \|_\infty$—specifically, for each $n \in \mathbf{N}$,

$$\|\hat{\mathbb{F}}_n - F\|_\infty \leq \|\mathbb{F}_n - F\|_\infty \ .$$

Kiefer and Wolfowitz (1976) show that $\hat{\mathbb{F}}_n$ in fact is asymptotically equivalent to $\mathbb{F}_n$ by essentially establishing the following result:

Theorem 1.1. *Let $\alpha_F \equiv \inf\{x \in \mathbf{R} : F(x) = 1\}$ and F be twice continuously differentiable on $[0, \alpha_F]$ with $\sup\{x \in \mathbf{R} : F(x) = 0\} = 0$. If $\alpha_F < \infty$ and*

$$\beta_F \equiv \inf_{0 < x < \alpha_F} \frac{-f'(x)}{f^2(x)} > 0 \ , \tag{1.1}$$

$$\gamma_F \equiv \frac{\sup_{0 < x < \alpha_F} -f'(x)}{\inf_{0 < x < \alpha_F} f^2(x)} < \infty \ , \tag{1.2}$$

then it follows that

$$\sup_{0 \leq x \leq \alpha_F} |\hat{\mathbb{F}}_n(x) - \mathbb{F}_n(x)| = O(n^{-2/3}(\log n)^{2/3}) \ \text{almost surely} \ .$$

Since $F(0) = 0$ by assumption, it follows from $\alpha_F < \infty$ that the distribution associated with F has bounded support. In turn, the condition (1.1) implies that $f'(x) < 0$ for all $x \in (0, \alpha_F)$ and hence that F is strictly concave on the support $[0, \alpha_F]$. Finally, the condition (1.2) demands that the derivative f' of the density be small relative to the minimum of f on its support. This paper provides refinements of Theorem 1.1 along several dimensions. First, we consider situations where f has potentially unbounded support, i.e., $\alpha_F = \infty$, an extension which appears to be nontrivial—see Remark 5 in Kiefer and Wolfowitz (1976, p.82) and also Kiefer and Wolfowitz (1977). In doing so, we necessarily have $\gamma_F = \infty$ (so that (1.2) is being violated) because the density f must approach zero along the right tail. Our second refinement, which may or may not be surprising, allows F to be non-strictly concave or even non-concave but requires that it be "asymptotically" strictly concave as the sample size tends to infinity. In any case, the above asymptotic order result is uncovered under weaker regularity conditions.

The second set of results concerns inference aspect of the problem. In particular, we develop a test for the hypothesis that F is concave or equivalently f is nonincreasing. The insight we exploit here is that the hypothesis can be equivalently formulated as:

$$\mathrm{H}_0 : \|\hat{F} - F\|_p = 0 \qquad \text{v.s.} \qquad \mathrm{H}_1 : \|\hat{F} - F\|_p > 0 \ ,$$

where $\hat{F}$ is the least concave majorant of F, and $\| \cdot \|_p$ is an L^p norm (with appropriate weighting) for $p \in [1, \infty]$. Thus, it is natural for us to employ the test statistic $\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_p$. There are several technical challenges, however, in establishing statistical properties of our test. First, as demonstrated in Beare and Moon (2015) and Beare and Fang (2017), the least concave majorant operator $F \mapsto \hat{F}$ is not fully differentiable if F is concave but not strictly concave,

and hence the conventional Delta method is inapplicable in establishing the asymptotic distribution of $\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_p$. However, the same authors have also shown that it is Hadamard directionally differentiable, a weaker notion of differentiability under which the Delta method is in fact preserved (Shapiro, 1991; Dümbgen, 1993). Second, as shown by Dümbgen (1993) and Fang and Santos (2019), even though the Delta method generalizes to deriving asymptotic distributions under Hadamard directional differentiability, it does not generalize to bootstrap consistency. This can be remedied by appealing to the rescaled bootstrap of Dümbgen (1993). Lastly, the weak limit of $\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_p$ is degenerate at zero when F is strictly concave, a consequence of the Kiefer-Wolfowitz theorem. To build up a level α test, we leverage the asymptotic order results on $\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty$ and propose a selection procedure we call the KW-selection that determines whether F is strictly concave or not under the null. The idea is that the convergence rates of $\hat{\mathbb{F}}_n$ are different in these two cases, and hence we may introduce a suitable tuning parameter to identify the truth. Our test is in fact shown to have local size control which may be of interest since the weak limit of the statistic exhibits "discontinuity" with respect to the true distribution. Unfortunately, the local power will be poor under local (contiguous) alternatives that approach a strictly concave distribution.

The literature on testing concavity of the cumulative distribution function (or equivalently monotonicity of the density function) is surprisingly limited. In Section 3.3, we review related studies in this regard and compare them to our test. Overall, there are roughly two types of tests in the density context. The first type is concerned with special classes of alternatives. This includes the work of Woodroofe and Sun (1999) who test uniformity against monotonically increasing (but not uniform) densities, or uniformity against convex cdfs. Therefore, these tests are powerful in detecting these special classes by design, but may perform poorly for other alternatives, which is confirmed in our simulation studies. The second type remains agnostic about the particular natures of alternative hypothesis, but relies on critical values from the asymptotically least favorable distributions. This includes the tests of Durot (2003) and Kulikov and Lopuhaä (2004). While the use of least favorable distributions is common practice in some other settings (Carolan and Tebbs, 2005; Delgado and Escanciano, 2012, 2013, 2016), the resulting tests may be too conservative and cause substantial power loss. Finally, in similar contexts, there are recent studies such as Seo (2018) and Beare and Shi (2019) that aim to improve power by bootstrap. However, due to the lack of Kiefer-Wolfowitz type results, it is unclear what happens to their tests when the limiting distributions of the test statistics are degenerate. We shall remove the dependence on least favorable distributions by bootstrap, and control the size of our test (both pointwise and locally) even when degeneracy occurs by utilizing the Kiefer-Wolfowitz's theorem.

We now introduce some notation. We denote by $\mathbf{R}_+$ the set of nonnegative real numbers. For a sequence $\{Z_n\}_{n=1}^\infty$ of random variables, we write $Z_n = O_{a.s.}(a_n)$ for a deterministic sequence $\{a_n\}_{n=1}^\infty$ if $Z_n/a_n = O(1)$ as $n \to \infty$ almost surely. Here and elsewhere in this paper, for a sequence $\{b_n\}$ of real numbers, $b_n = O(1)$ as $n \to \infty$ means that $\{b_n\}$ is bounded. For two sequences

of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ if $a_n/b_n = O(1)$ as $n \to \infty$, and write $a_n \gtrsim b_n$ if $b_n \lesssim a_n$. Moreover, the notation $a_n \asymp b_n$ means that $a_n/b_n = O(1)$ and $b_n/a_n = O(1)$ as $n \to \infty$. Analogously, for two functions $f, g :$ $\mathbf{R} \to \mathbf{R}_+$, we write $f(\epsilon) \asymp g(\epsilon)$ as $\epsilon \downarrow 0$ if $f(\epsilon)/g(\epsilon) = O(1)$ and $g(\epsilon)/f(\epsilon) = O(1)$ as $\epsilon \downarrow 0$. For an arbitrary nonempty set T, $\ell^\infty(T)$ is the space of bounded real-valued functions on $\mathbf{R}_+$ equipped with the uniform norm $\| \cdot \|_\infty$, i.e., $\|f\|_\infty \equiv \sup_{x \in T} |f(x)|$. The space $C_0(\mathbf{R}_+)$ is the family of all real-valued continuous functions on $\mathbf{R}_+$ that vanish at infinity. We denote by $\xrightarrow{L}$ the weak convergence in the sense of Hoffmann-Jørgensen (van der Vaart and Wellner, 1996).

The remainder of the paper is structured as follows. Section 2 presents refinements of Theorem 1.1. Section 3 formally develops our test and then compare to some existing monotonicity tests from the literature. Section 4 conducts simulation studies. Section 5 concludes. All proofs are collected in Appendix A, while extensions of the test results to a general setup are discussed in Appendix B. Additional simulation results may be found in Appendix C of Fang (2019).

2. The Kiefer-Wolfowitz theorems

This section presents refinements of the Kiefer-Wolfowitz theorem that allow the support of the density function f to be unbounded. Throughout, we think of the cdf F as a function on $\mathbf{R}_+$ (rather than on $\mathbf{R}$). We proceed with the following assumption.

Assumption 2.1. (i) $\{X_i\}_{i=1}^n$ is an i.i.d. sample with common distribution function F, and (ii) F is strictly concave on $\mathbf{R}_+$.

Assumption 2.1 simply formalizes the i.i.d. setup and strict concavity of the distribution function F. To establish the refinements, we need to impose further regularity conditions on F. In particular, we need to introduce an analog of β_F in (1.1). Specifically, for any $\epsilon \in [0, 1]$, we define

$$\bar{\beta}(\epsilon) \equiv \inf_{0 \le x \le F^{-1}(1-\epsilon)} \frac{-f'(x)}{f^2(x)} \, , \qquad (2.1)$$

where we suppress the dependence of $\bar{\beta}(\epsilon)$ on the underling distribution for simplicity. Thus, $\bar{\beta}(\epsilon)$ may be viewed as a truncated version of β_F in (1.1). As it turns out, while $\bar{\beta}(\epsilon)$ is allowed to approach zero as $\epsilon \to 0$, the speed at which it approaches zero determines the asymptotic order of $\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty$. To formalize our discussions, we now impose:

Assumption 2.2. (i) F is twice continuously differentiable on $\mathbf{R}_+$, (ii) $\bar{\beta}(\epsilon) \asymp \epsilon^\tau$ as $\epsilon \to 0$ for some constant $\tau > -1$, and (iii)

$$\|F''\|_{1/2} \equiv \left[\int_0^\infty \sqrt{|F''(x)|} \, \mathrm{d}x \right]^2 < \infty \, .$$

Assumption 2.2(i) is the same smoothness condition required by Theorem 1.1. Assumption 2.2(ii) characterizes the exact rates at which $\bar\beta(\epsilon)$ is allowed to approach zero as $\epsilon \to 0$. The special case $\tau = 0$ implies $\inf_{0 \le x < \infty} \{-f'(x)/f^2(x)\} > 0$, which is exactly the condition (1.1) in the case of unbounded support. If $\tau > 0$, then $\bar\beta(\epsilon) \to 0$ as $\epsilon \to 0$; if $\tau \in (-1, 0)$, then $\bar\beta(\epsilon) \to \infty$ as $\epsilon \to 0$ which still fulfills (1.1). Assumption 2.2(iii) is a technical condition that serves to control the interpolation error for the cdf F. When the support of the density function f is bounded, Assumption 2.2(iii) is automatically fulfilled because F'' as a continuous function is bounded. Thus, our assumptions are strictly weaker than those imposed in Kiefer and Wolfowitz (1976).

Given Assumptions 2.1 and 2.2, we now present the first refinement.

Theorem 2.1. *If Assumptions 2.1 and 2.2 hold, then, as $n \to \infty$,*

$$\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty = \begin{cases} O_{a.s.}\big(((\log n)/n)^{\frac{2}{2\tau+3}}\big) & \text{if } \tau \ge 0 \\ O_{a.s.}\big(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}\big) & \text{if } \tau \in (-1, 0) \end{cases} .$$

Theorem 2.1 delivers the asymptotic order of $\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty$, which crucially depends on the parameter τ. The slower $\bar\beta(\epsilon)$ approaches zero (as $\epsilon \to 0$) or the smaller τ is, the closer $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ is (asymptotically). If $\bar\beta(\epsilon)$ approaches zero too fast or τ is too large, then the asymptotic equivalence of $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ are no longer implied. This happens precisely when $\tau \ge 1/2$, in which case

$$\sqrt{n}((\log n)/n)^{\frac{2}{2\tau+3}} = n^{\frac{2\tau-1}{2\tau+3}}(\log n)^{\frac{2}{2\tau+3}} \to \infty \text{ as } n \to \infty . \tag{2.2}$$

The special case $\tau = 0$ leads to $\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty = O_{a.s.}\big(((\log n)/n)^{2/3}\big)$, which is exactly the result in Theorem 1.1. There are two major new ingredients in establishing the theorem, compared to the proof of Theorem 1.1. One is controlling the variation of $\mathbb{F}_n$ over a small upper quantile region—see the treatment of the T_{n,k_n+1} term and the B_n term in the proof. The other is bounding the interpolation error for F by the $L_{1/2}$-integral of F'' based on a result from approximation theory (Burchard, 1974). The latter allows us to dispense with the regularity condition $\gamma_F < \infty$ (Kiefer and Wolfowitz, 1976, p.82).

Our second refinement allows the common distribution function F that the finite sample $X_1, \ldots, X_n$ share to be not strictly concave or even non-concave. However, we do require that it vary with the sample size n in such a way that it approaches (in a suitable sense to be specified) a strictly concave distribution function as $n \to \infty$, a notion we call "local to concavity." In order to formalize our local analysis, we denote by $\mathbf{P}$ the set of probability measures with the support contained in $\mathbf{R}_+$ that possibly govern the data:

$$\mathbf{P} \equiv \{P \in \mathbf{M} : P(\mathbf{R}_+) = 1\} ,$$

where $\mathbf{M}$ is the set of all Borel probability measures on $\mathbf{R}$ that are dominated by the Lebesgue measure on $\mathbf{R}$. Further, we think of the distribution function $F \equiv F(P)$ as a map $F : \mathbf{P} \to \ell^\infty(\mathbf{R}_+)$ defined by

$$F(P)(x) \equiv P((-\infty, x]) , \, \forall\, x \in \mathbf{R}_+ .$$

We may now formalize the precise meaning of "local" as follows.

Definition 2.1. A function $t \mapsto P_t$ mapping a neighborhood $(-\epsilon, \epsilon)$ of zero into $\mathbf{P}$ is called a differentiable path passing through P if, for $P_0 = P$ and some function $h : \mathbf{R}_+ \to \mathbf{R}$,

$$\lim_{t \to 0} \int \left[\frac{\mathrm{d}P_t^{1/2} - \mathrm{d}P^{1/2}}{t} - \frac{1}{2} h \, \mathrm{d}P^{1/2} \right]^2 = 0 \ . \tag{2.3}$$

The notation $\mathrm{d}P_t$ and $\mathrm{d}P$ may be understood as the densities of P_t and P with respect to some dominating measure μ_t (for each t), though the integral in (2.3) does not depend on the choice of μ_t (van der Vaart, 1998, p.362). Loosely speaking, (2.3) implies that P_t gets closer and closer to P_0 "on average", as $t \to 0$. The function h is referred to as the score function of P and satisfies $\int h \, \mathrm{d}P = 0$ and $h \in L^2(P)$—see, for example, Lemma 25.14 in van der Vaart (1998). The term "score function" makes sense in view of the relation $h(x) = \frac{\mathrm{d}}{\mathrm{d}t} \log \mathrm{d}P_t(x)\big|_{t=0}$ (which is the usual definition of score function) under regularity conditions—see, for example, Lemma 7.6 in van der Vaart (1998).

If $\{P_t : |t| < \epsilon\}$ is a differentiable path, then we obtain by Example 5.3.1 in Bickel et al. (1998) that

$$\lim_{t \to 0} \| \frac{F(P_t) - F(P)}{t} - \dot{F}(h) \|_\infty = 0 \ , \tag{2.4}$$

where

$$\dot{F}(h)(x) \equiv \int_{-\infty}^{x} h(u) \, P(\mathrm{d}u) \ .$$

Intuitively, (2.4) means that the distribution function $F_t \equiv F(P_t)$ smoothly passes through $F \equiv F(P_0)$ at the same speed as the underlying probability measure P_t passes through P_0. For a differentiable path $\{P_t\}$ satisfying (2.3) and $F_t \equiv F(P_t)$, we say $\{F_t\}$ is *local to concavity* if F is concave and is *local to strict concavity* if F is strictly concave.

We next show that the conclusion in Theorem 1.1 is preserved under local to strict concavity. To this end, we impose the following assumption.

Assumption 2.3. (i) $\{X_i\}_{i=1}^{n}$ is an i.i.d. sample with common probability measure $P_{1/\sqrt{n}}$, and (ii) $\{P_{1/\sqrt{n}}\} \subset \mathbf{P}$ corresponds to the differentiable path $\{P_t\}$ passing through $P_0 \equiv P$ in the sense of Definition 2.1.

Theorem 2.2. *If Assumptions 2.1(ii), 2.2 and 2.3 hold, then, as $n \to \infty$,*

$$\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty = \begin{cases} O_{a.s.}\big(((\log n)/n)^{\frac{2}{2\tau+3}}\big) & \text{if } \tau \ge 0 \\ O_{a.s.}\big(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}\big) & \text{if } \tau \in (-1, 0) \end{cases} \ .$$

Theorem 2.2 is not much deeper than Theorem 1.1, but it has implications for the problem of testing concavity as we shall elaborate in Section 3.

3. Testing concavity

In this section, we develop a test of concavity that controls size regardless of whether concavity is strict or not. This is accomplished by building on the asymptotic order results established previously. We shall also compare our concavity test with existing monotonicity tests. For simplicity, in what follows, we focus on the canonical case $\tau = 0$ in Assumption 2.2(ii). The general case makes no essential differences—see Appendix B for extensions.

3.1. The test statistic

First, we introduce our test statistic and derive its asymptotic distribution. To this end, we introduce the least concave majorant (LCM) operator following Beare and Fang (2017)—see also Beare and Moon (2015).

Definition 3.1. Given a convex set $T \subseteq \mathbf{R}_+$, the LCM over T is the operator $\mathcal{M}_T : \ell^\infty(\mathbf{R}_+) \to \ell^\infty(T)$ that maps each $\theta \in \ell^\infty(\mathbf{R}_+)$ to the function

$$\mathcal{M}_T\theta(x) \equiv \inf\{g(x) : g \in \ell^\infty(T), \, g \text{ is concave, and } \theta \leq g \text{ on } T\}, \quad \forall \, x \in T \ .$$

If $T = \mathbf{R}_+$, then we write $\mathcal{M} \equiv \mathcal{M}_{\mathbf{R}_+}$.

The hypothesis of our interest can now be formulated as:

$$\mathrm{H}_0 : \phi(F) = 0 \quad \text{v.s.} \quad \mathrm{H}_1 : \phi(F) > 0 \ ,$$

where $\phi : \ell^\infty(\mathbf{R}_+) \to \mathbf{R}$ is defined by

$$\phi(F) = \|\mathcal{M}F - F\|_p = \begin{cases} [\int_{\mathbf{R}_+} (\mathcal{M}F(x) - F(x))^p g(x) \, \mathrm{d}x]^{1/p} & \text{if } p \in [1, \infty) \\ \sup_{x \in \mathbf{R}_+} |\mathcal{M}F(x) - F(x)| & \text{if } p = \infty \end{cases} \ ,$$

with $g \in L^1(\mathbf{R}_+)$ some known positive weighting function. Here and throughout, we work with an arbitrarily fixed $p \in [1, \infty]$, and thus suppress the dependence of ϕ on p for notational simplicity. Given the above formulation, we employ the statistic $\sqrt{n}\phi(\mathbb{F}_n)$. To derive the weak limit of $\sqrt{n}\phi(\mathbb{F}_n)$, note that under the null hypothesis, $\mathcal{M}F = F$ and hence we may rewrite:

$$\sqrt{n}\phi(\mathbb{F}_n) = \sqrt{n}\|\mathcal{M}\mathbb{F}_n - \mathbb{F}_n\|_p = \|\sqrt{n}\{\mathcal{D}\mathbb{F}_n - \mathcal{D}F\}\|_p \ , \tag{3.1}$$

where $\mathcal{D} \equiv \mathcal{M} - \mathcal{I}$ with $\mathcal{I}$ the identity operator on $\ell^\infty(\mathbf{R}_+)$. Thus, the asymptotic distribution of $\sqrt{n}\phi(\mathbb{F}_n)$ would be an implication of the continuous mapping theorem and the Delta method, if we could show that $\mathcal{M}$ (and hence $\mathcal{D}$) is Hadamard differentiable (van der Vaart and Wellner, 1996, p.372-374). Unfortunately, as demonstrated by Beare and Moon (2015) and Beare and Fang (2017), the LCM operator fails to be fully differentiable but only Hadamard directionally differentiable in general (Shapiro, 1990). Nonetheless, this type of directional differentiability suffices for applying a generalized version of the Delta method (Shapiro, 1991; Dümbgen, 1993).

Definition 3.2. Let $\mathbb{D}$ and $\mathbb{E}$ be normed spaces equipped with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$ respectively, and $\phi : \mathbb{D}_{\phi} \subseteq \mathbb{D} \to \mathbb{E}$. The map ϕ is said to be *Hadamard directionally differentiable* at $\theta \in \mathbb{D}_{\phi}$ *tangentially* to $\mathbb{D}_0 \subseteq \mathbb{D}$, if there is a map $\phi'_{\theta} : \mathbb{D}_0 \to \mathbb{E}$ such that:

$$\lim_{n \to \infty} \|\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_{\theta}(h)\|_{\mathbb{E}} = 0 , \qquad (3.2)$$

for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \subset \mathbf{R}_+$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{D}_0$ as $n \to \infty$ and $\theta + t_n h_n \in \mathbb{D}_{\phi}$ for all n.

The defining feature of Hadamard directional differentiability is that, unlike Hadamard (full) differentiability, the directional derivative is in general non-linear though necessarily continuous and positively homogeneous of degree one (Shapiro, 1990).[1] We refer the reader to Shapiro (1990) and a more recent review by Fang and Santos (2019) for additional discussions. Proposition 2.1 in Beare and Fang (2017) implies that $\mathcal{M}$ is Hadamard directionally differentiable at any concave $F \in \ell^{\infty}(\mathbf{R}_+)$ tangentially to the set $C_0(\mathbf{R}_+)$ with the derivative $\mathcal{M}'_F : C_0(\mathbf{R}_+) \to \ell^{\infty}(\mathbf{R}_+)$ given by: for any $h \in C_0(\mathbf{R}_+)$ and $x \in \mathbf{R}_+$,

$$(\mathcal{M}'_F h)(x) = (\mathcal{M}_{T_{F,x}} h)(x) ,$$

where $T_{F,x} = \{x\} \cup U_{F,x}$ with $U_{F,x}$ the union of all open intervals $A \subseteq \mathbf{R}_+$ such that (i) $x \in A$, and (ii) F is affine over A. We emphasize that $\mathcal{M}'_F$ is equal to the identity operator on $C_0(\mathbf{R}_+)$ if and only if F is strictly concave, in which case $\mathcal{M}$ is Hadamard differentiable at F tangentially to $C_0(\mathbf{R}_+)$—see Proposition 2.2 in Beare and Fang (2017) for more details.

We are now in a position to state the asymptotic distribution of our statistic $\sqrt{n}\phi(\mathbb{F}_n)$ by invoking the generalized Delta method.

Lemma 3.1. *If Assumption 2.1(i) holds, then $\sqrt{n}\phi(\mathbb{F}_n) \xrightarrow{L} \|\mathcal{M}'_F \mathbb{G} - \mathbb{G}\|_p$ under* H_0, *where $\mathbb{G} \equiv \mathbb{B} \circ F$ and $\mathbb{B}$ is the standard Brownian bridge on $[0,1]$.*

Lemma 3.1 establishes the weak limit of the test statistic $\sqrt{n}\phi(\mathbb{F}_n)$ under the null. The Brownian bridge $t \mapsto \mathbb{G}(t) \equiv \mathbb{B}(F(t))$ is a zero mean Gaussian process with covariance function: for all $s, t \in [0,1]$,

$$\mathrm{Cov}(\mathbb{G}(t), \mathbb{G}(s)) = F(\min(s,t)) - F(t)F(s) . \qquad (3.3)$$

The limiting distribution in Lemma 3.1 is not pivotal because it depends on F through the process $\mathbb{G}$ and, critically, the derivative $\mathcal{M}'_F$.

3.2. The critical values

Towards constructing critical values for our test, we next aim to estimate the law of $\|\mathcal{M}'_F \mathbb{G} - \mathbb{G}\|_p$ through bootstrap. There are, however, two complications involved, as we now elaborate.

---

[1] Let $\mathbb{D}$ be a vector space. A function $f : \mathbb{D} \to \mathbf{R}$ is said to be positively homogeneous of degree one if and only if $f(ax) = af(x)$ for all $x \in \mathbb{D}$ and all $a \in \mathbf{R}_+$.

4604 Z. Fang

First, when F is non-strictly concave in which case $\mathcal{M}$ is only Hadamard directionally differentiable, the standard bootstrap (compare to (3.1)),

$$\|\sqrt{n}\{\mathcal{D}\mathbb{F}_n^* - \mathcal{D}\mathbb{F}_n\}\|_p \ , \tag{3.4}$$

is necessarily inconsistent. This is a consequence of Proposition 1 in Dümbgen (1993), which has been formalized by Theorem A.1 in Fang and Santos (2019). Second, when F is strictly concave, the weak limit of $\sqrt{n}\phi(\mathbb{F}_n)$ is degenerate at zero since $\mathcal{M}'_F = \mathcal{I}$, which is not surprising in view of Theorem 2.1. The first issue can be resolved by appealing to the rescaled bootstrap in Dümbgen (1993). As shall be seen shortly, the rescaled bootstrap is connected in a subtle way to the modified bootstrap in Fang and Santos (2019); namely, it amounts to composing a suitable derivative estimator (see (3.7)) with some bootstrap process. The second issue is more challenging, which we shall fix by leveraging Theorems 2.1 and 2.2 as we describe now.

The key to the level control of our test is a selection procedure that determines whether the concavity of F is strict or not by exploiting the fact that the convergence rates in these two cases are different (see Theorem 2.1 and Lemma 3.1). Specifically, let $\{\kappa_n\}$ be a sequence of positive scalars such that $\kappa_n = o(1)$ and $(\log n)^{2/3}n^{-1/6}/\kappa_n = o(1)$ as $n \to \infty$. For example, we may take $\kappa_n = n^{-1/7}$ or $(\log n)^{-1}$. Define for each $n \in \mathbf{N}$,

$$\xi_n \equiv \frac{\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_p}{\kappa_n} \ . \tag{3.5}$$

Then $\xi_n \xrightarrow{p} 0$ if F is strictly concave by Theorem 2.1 and $\xi_n \xrightarrow{p} \infty$ if F is non-strictly concave by Lemma 3.1. Thus, if $\sqrt{n}\phi(\mathbb{F}_n) \le \kappa_n$, we may take F to be strictly concave and non-strictly concave otherwise. We refer to such a selection procedure as the KW-selection. As a result, the asymptotic level of our test can be controlled in this case by choosing the critical value to be κ_n.

On the other hand, when F is non-strictly concave, the nondegenerate limit $\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p$ in Lemma 3.1 can be estimated by composing a suitable estimator $\hat{\mathcal{M}}'_n$ of $\mathcal{M}'_F$ with the nonparametric bootstrap estimator $\sqrt{n}\{\mathbb{F}_n^* - \mathbb{F}_n\}$ of the Gaussian process $\mathbb{G}$, i.e., we employ the bootstrap estimator

$$\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}_n^* - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}_n^* - \mathbb{F}_n\}\|_p \ , \tag{3.6}$$

where $\mathbb{F}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{ni}1\{X_i \le x\}$ for $x \in \mathbf{R}_+$, $W_n \equiv (W_{n1}, \ldots, W_{nn})$ a multinomial random vector with n categories and probabilities $(1/n, \ldots, 1/n)$, and $\hat{\mathcal{M}}'_n : \ell^\infty(\mathbf{R}_+) \to \ell^\infty(\mathbf{R}_+)$ is some appropriate derivative estimator. Such a bootstrap procedure is justified by Theorem 3.2 in Fang and Santos (2019). It is in fact precisely the rescaled bootstrap proposed by Dümbgen (1993), which amounts to estimating $\mathcal{M}'_F$ by: for any $h \in \ell^\infty(\mathbf{R}_+)$,

$$\hat{\mathcal{M}}'_n(h) = \frac{\mathcal{M}(\mathbb{F}_n + t_n h) - \mathcal{M}(\mathbb{F}_n)}{t_n} \ , \tag{3.7}$$

where $t_n \downarrow 0$ such that $t_n\sqrt{n} \to \infty$—see Fang and Santos (2019, p.390-1). In turn, we may choose the critical value in this case as: for $\alpha \in (0,1)$,

$$\hat{c}^*_{1-\alpha} \equiv \inf\{c \in \mathbf{R} : P_W(\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p \le c) \ge 1 - \alpha\} \ ,$$

where P_W denotes the probability with respect to the bootstrap weights $\{W_{ni}\}^n_{i=1}$ holding the data $\{X_i\}^n_{i=1}$ fixed. In practice, $\hat{c}^*_{1-\alpha}$ can be estimated by Monte Carlo simulations by drawing a large number of bootstrap samples $\{X^*_{b,i}\}^n_{i=1}$ from $\mathbb{F}_n$ for $b = 1, \ldots, B$ (with the data $\{X_i\}^n_{i=1}$ fixed).

Now, for each fixed $\alpha \in (0,1)$ we set the critical value of our test as

$$\hat{c}_{1-\alpha} \equiv \max\{\kappa_n, \hat{c}^*_{1-\alpha}\} \ ,$$

which provides pointwise asymptotic level control as confirmed by the following theorem. Before stating the theorem, we formalize requirements on t_n, κ_n, and the $(1 - \alpha)$th quantile $c^*_{1-\alpha}$ of the cdf of the weak limit $\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p$.

Assumption 3.1. (i) $\{t_n\}$ is a sequence of positive scalars such that $t_n \downarrow 0$ such that $t_n\sqrt{n} \to \infty$ as $n \to \infty$; (ii) $\{\kappa_n\}$ is a sequence of positive scalars such that $\kappa_n \downarrow 0$ and $(\log n)^{2/3}n^{-1/6}/\kappa_n \to 0$ as $n \to \infty$.

Assumption 3.2. The cdf of $\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p$ is continuous and strictly increasing at its $(1 - \alpha)$th quantile $c^*_{1-\alpha}$ when F is non-strictly concave.

Assumption 3.1 imposes the convergence rates on the tuning parameters κ_n and t_n. We do not touch the challenging issue of optimal choices in this paper. Assumption 3.2 is a standard technical condition (often implicitly imposed) to ensure that consistent bootstrap can produce consistent critical values—see Lemma 11.2.1 in Lehmann and Romano (2005). In verifying Assumption 3.2, we note that, by the proof of Theorem 3.2, the map $g \mapsto \phi'_F(g) \equiv \|\mathcal{M}'_F g - g\|_p$ is subadditive[2] and hence convex since it is obviously positively homogeneous of degree one. Hence, by Theorem 11.1 in Davydov, Lifshits and Smorodina (1998), the distribution function H of $\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p$ is absolutely continuous and strictly increasing on (r_0, ∞), with $r_0 \equiv \inf\{r \in \mathbf{R} : H(r) > 0\}$. Therefore, Assumption 3.2 holds whenever $c^*_{1-\alpha} > r_0$.

We are now in a position to state the first main result of this paper.

Theorem 3.1. *If Assumptions 2.1(i), 2.2 with $\tau = 0$, 3.1 and 3.2 hold, then it follows that, under* H_0,

$$\limsup_{n\to\infty} P(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{1-\alpha}) \le \alpha \ , \tag{3.8}$$

and under H_1,

$$\lim_{n\to\infty} P(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{1-\alpha}) = 1 \ . \tag{3.9}$$

---

[2]Let $\mathbb{D}$ be a vector space. A function $f : \mathbb{D} \to \mathbf{R}$ is said to be subadditive if and only if $f(x + y) \le f(x) + f(y)$ for all $x, y \in \mathbb{D}$.

Theorem 3.1 shows that our test is pointwise (in P) asymptotically level α and consistent under any fixed alternative. However, in view of the irregularity of the problem and as argued in the literature (Lehmann and Romano, 2005; Andrews and Guggenberger, 2010; Romano and Shaikh, 2012), pointwise asymptotics can be unreliable in "nonstandard" settings. It is therefore of interest to investigate the uniform or at least local properties of our test.

To study both local size control and local power of our proposed test, we follow van der Vaart (1998, p.384-6) and consider differentiable paths in $\mathbf{P}$ passing through $P_0 \equiv P$ that also belong to the set

$$\mathbf{H} \equiv \{\{P_t\} : (\text{i})\phi(F(P_t)) = 0 \text{ if } t \leq 0, \text{ and } (\text{ii})\phi(F(P_t)) > 0 \text{ if } t > 0\} .$$

Thus, a differentiable path $\{P_t\}$ in $\mathbf{H}$ is such that if it satisfies the null hypothesis whenever $t \leq 0$ but otherwise the alternative for all $t > 0$. For a differentiable path $\{P_{\eta/\sqrt{n}}\}$ where $\eta \in \mathbf{R}$, we set $P^n \equiv \prod_{i=1}^n P$, $P_n^n \equiv \prod_{i=1}^n P_{\eta/\sqrt{n}}$ and define the power function of our test for sample size n as

$$\pi_n(P_{\eta/\sqrt{n}}) \equiv P_n^n(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{1-\alpha}) .$$

Our next theorem establishes local properties of our test.

Theorem 3.2. *Let $\{P_t\}$ be a differentiable path in $\mathbf{H}$ and let Assumptions 2.2, 2.3, 3.1, and 3.2 hold with $\tau = 0$ and $F \equiv F(P_0)$. Then it follows that*

1. *For any $\eta \in \mathbf{R}$, (i) if F is strictly concave, then $\liminf_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) = 0$, and (ii) if F is non-strictly concave, then*

$$\liminf_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}})$$
$$\geq P(\|\mathcal{M}'_F(\mathbb{G} + \eta\dot{F}(h)) - \{\mathbb{G} + \eta\dot{F}(h)\}\|_p > c^*_{1-\alpha}) , \quad (3.10)$$

   where (3.10) holds with equality if, in addition, the cdf of $\|\mathcal{M}'_F(\mathbb{G} + \eta\dot{F}(h)) - \{\mathbb{G} + \eta\dot{F}(h)\}\|_p$ is continuous at $c^*_{1-\alpha}$.
2. *For any $\eta \leq 0$, we have*

$$\limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) \leq \alpha . \quad (3.11)$$

3. *If $\|\mathcal{M}'_F(\dot{F}(h)) - \dot{F}(h)\|_p > 0$, then*

$$\liminf_{\eta\uparrow\infty} \liminf_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) = 1 . \quad (3.12)$$

The first part of Theorem 3.2 delivers a lower bound of the local limiting power function. If the true distribution function $F_n \equiv F(P_{\eta/\sqrt{n}})$ is local to a strictly concave function F, then the limiting local power along the path $\{P_{1/\sqrt{n}}\}$ is zero. This is unfortunate and in fact is an implication of Theorem 2.2. The second part shows that the asymptotic null rejection rate along $\{P_{1/\sqrt{n}}\}$ is no larger than α, establishing the asymptotic local size control.

The first part of Theorem 3.2 might leave one the impression that the test has poor local power against any sequence of local alternatives $\{P_{\eta/\sqrt{n}}\}$ (with $\eta > 0$).

The third part is intended to reconcile such a misconception. Heuristically, it says that our test has nontrivial local power if the sequence $\{P_{\eta/\sqrt{n}}\}$ does not approach the null too fast. To appreciate the condition $\|\mathcal{M}'_F(\dot{F}(h)) - \dot{F}(h)\|_p > 0$, we first note that it prevents F from being strictly concave in which case $\mathcal{M}'_F$ is the identity operator and the limiting local power is zero by Part 1-(i). Second, by (2.4), we have that, as $n \to \infty$,

$$\sqrt{n}\{F_n - F\} = \sqrt{n}\{F(P_{h/\sqrt{n}}) - F(P)\} \to \eta \dot{F}(h) \ . \tag{3.13}$$

In turn, this implies by Proposition 2.1 in Beare and Fang (2017) and $\phi(F) = 0$ (by the definition of **H**) that, as $n \to \infty$,

$$\sqrt{n}\phi(F_n) = \sqrt{n}\{\phi(F_n) - \phi(F)\} \to \|\mathcal{M}'_F(\eta \dot{F}(h)) - \eta \dot{F}(h)\|_p \ . \tag{3.14}$$

By the positive homogeneity of degree one of $\mathcal{M}'_F$ (as a Hadamard directional derivative), we thus have: for $\Delta \equiv \|\mathcal{M}'_F(\dot{F}(h)) - \dot{F}(h)\|_p$,

$$\phi(F_n) = \frac{\eta}{\sqrt{n}}(\Delta + o(1)) \ , \tag{3.15}$$

for all $\eta > 0$. Therefore, $\Delta \equiv \|\mathcal{M}'_F(\dot{F}(h)) - \dot{F}(h)\|_p > 0$ implies that the third part is concerned with local power along a nontrivial Pitman drift, a canonical device for local power analysis. When $\Delta = 0$ (as when F is strictly concave), the speed that $\{P_{\eta/\sqrt{n}}\}$ approaches the null is too fast in the sense that $\sqrt{n}\phi(F_n) \to 0$, thereby making the test hard to reject.

3.3. Comparisons with existing tests

We now compare our concavity test with some existing tests. While there is a rich literature on testing monotonicity in settings such as nonparametric regression—see Chetverikov (2019) for a recent study with a brief survey, it is somewhat surprising that results for the same problem in the density context seem rather limited. This is just one piece of evidence consistent with Jon Wellner's view that "the (shape-constrained) community also needs to do more work to provide inferential methods beyond estimation" (Banerjee and Samworth, 2018).

Banerjee and Wellner (2001) propose likelihood ratio tests for equality at a fixed point, rather than the global shape as we consider, in the setting of nonparametric estimation of a monotone function—see also Banerjee (2005). More related to our setting is the work by Woodroofe and Sun (1999), who are concerned with testing uniformity versus a monotone (but not uniform) density. Translated to our setup, they study the simple null F being the uniform distribution against the composite alternative that F is convex (but is not a uniform distribution). Therefore, the parameter spaces under the null and the alternative are smaller than the corresponding spaces in our setup. Woodroofe and Sun (1999) propose two tests, namely, the P-test and the D-test, based on a penalized nonparametric maximum likelihood estimator of the density. The P-test statistic is simply the penalized likelihood ratio, while the D-test statistic

is $\sqrt{n}$—with n being the sample size—times the uniform distance between the penalized cdf estimator and the uniform cdf.

The simple null hypothesis in Woodroofe and Sun (1999), though greatly simplifies the asymptotic analysis, may be restrictive from a practical point of view. This is a reflection of the fact that the pointwise asymptotic distributions of the test statistic under a composite null such as ours are highly nonstandard. In a nonparametric regression setting, Durot (2003) studies the composite null that the regression function is nonincreasing against the alternative that it is not. The test statistic proposed by Durot is a rescaled version of the uniform distance between a cumulative regression estimator and its least concave majorant. Durot (2003) shows that the constant regression functions are asymptotically least favorable, base on which critical values are constructed. Kulikov and Lopuhaä (2004) adapt Durot (2003)'s test to the density setup based on the same critical values—see their Theorem 3.1, and also Proposition 3.1 in Kulikov and Lopuhaä (2008). Kulikov and Lopuhaä (2004) also propose a test based on the L_p-norm (against the empirical distribution) of the empirical cdf and its least concave majorant, also relying on critical values constructed from the least favorable distribution. We note that the test based on the statistic T_n in Kulikov and Lopuhaä (2004) is invalid in the sense that it may over-reject under the null, as pointed out by the same authors and confirmed by their simulations.

Tests based on the least favorable distributions, though control the size in a very simple way, may be too conservative and result in power loss. This includes the tests of Durot (2003) and Kulikov and Lopuhaä (2004). On the other hand, the P-test and the D-test explicitly take into account the alternative hypothesis of nondecreasing densities. Therefore, they are powerful in detecting nondecreasing densities, but may perform poorly when the underlying distribution is neither concave nor convex. Our simulations confirm these predictions. Finally, we reiterate that the need of bootstrap for testing concavity as well as the tuning parameters κ_n and t_n is in line with the nonstandard nature of the problem, rather than a special attribute of our inferential framework. The implementation of our bootstrap is as simple as calculating the test statistic, or as simple as computing the least favorable majorant.

4. Simulation experiments

We next evaluate the finite sample performance of our test through Monte Carlo simulations. Special attention shall be paid to the tuning parameters κ_n (for the KW selection) and t_n (for the rescaled bootstrap). For this, we consider $\kappa_n \in \{n^{-1/7}, n^{-1/8}, \log^{-1} n\}$ and $t_n \in \{n^{-1/\varpi} : \varpi = 3, 4, \ldots, 7\}$. We run two sets of simulations in order to compare with the standard bootstrap and some existing monotonicity tests. Throughout, we let the significance level be 5%, and all results are based on 1000 Monte Carlo simulations and 500 bootstrap repetitions for each simulation replication (to implement our test).

First, we consider the distributions defined by the following densities supported on $\mathbf{R}_+$: for any $x \in \mathbf{R}_+$,

$$f_1(x) = \frac{1}{\sqrt{\pi/2}} \exp\{-\frac{x^2}{2}\} , \tag{4.1}$$

$$f_2(x) = \begin{cases} 3/2 & \text{if } 0 \leq x < 1/2 \\ 1/2 & \text{if } 1/2 \leq x < 9/10 \\ \exp\{9/10 - x\}/20 & \text{if } x \geq 9/10 \end{cases} . \tag{4.2}$$

The function f_1 is strictly decreasing, while f_2 is weakly decreasing though strictly decreasing on the region $[9/10, \infty)$. Hence, the corresponding distribution functions, denoted F_1 and F_2, are strictly concave and non-strictly concave respectively. To examine local behaviors of our test, we construct differentiable paths passing through F_1 and F_2 as follows: for $t \in \mathbf{R}_+$ and $x \in \mathbf{R}_+$,

$$f_{1,t}(x) = \frac{\exp\{-tx\}f_1(x)}{\int_{\mathbf{R}_+} \exp\{-tx\}f_1(x)dx} = \frac{1}{a(t)} \exp\{-\frac{(x+t)^2}{2}\} , \tag{4.3}$$

$$f_{2,t}(x) = \begin{cases} e^{-t}3/2 & \text{if } 0 \leq x < 1/2 \\ e^t/2 & \text{if } 1/2 \leq x < 9/10 \\ b(t)\exp\{9/10 - x\}/20 & \text{if } x \geq 9/10 \end{cases} , \tag{4.4}$$

where $a(t) \equiv \int_t^\infty \exp\{-u^2/2\}du$ and $b(t) \equiv 20 - 15e^{-t} - 4e^t$. By Example 3.2.1 and Proposition 2.1.1 in Bickel et al. (1998), these two paths are indeed differentiable (under the null). We then generate i.i.d. samples $\{X_i\}_{i=1}^n$ from (4.1), (4.2), (4.3) and (4.4), with $n \in \{100, 200, 300, 400, 1000\}$. For simplicity, we only consider the sup statistic $\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty$, and report results based on the rescaled and the standard bootstrap.

Tables 1, 2, 3 and 4 present the numerical results. Inconsistency of the standard bootstrap is prominently evidenced in last columns of Tables 2 and 4 which are based on non-strictly concave functions (so that the operator $\mathcal{M}$ are only Hadamard directionally differentiable at these distribution functions). On the contrary, our test alleviates the size distortion both pointwise and locally, though the results vary with the choices of the tuning parameters. Overall, the choices $t_n \in \{n^{-1/3}, n^{-1/4}\}$ tend to be too small, while $t_n \in \{n^{-1/5}, n^{-1/6}, n^{-1/7}\}$ are adequate in that they control the size well. Tables 1 and 3 also show that, when the distribution function is strictly concave, our test by design controls the size both pointwise and locally, though the rejection rates are often close to zero, which is consistent with Theorem 3.2. Moreover, as in Tables 1 and 3, the standard bootstrap also exhibits size control, due to the facts that the LCM operator is fully (Hadamard) differentiable at strictly concave distribution functions and that the distance between the Grenander distribution estimator and the empirical cdf converges faster than $\sqrt{n}$ by the Kiefer-Wolfowitz theorems. Finally, we note that the results are quite insensitive to the choice of κ_n.

TABLE 1
Pointwise Size Control with Strict Concavity Based on (4.1)

		Rescaled					
				t_n			Standard
n	κ_n	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	
	$n^{-1/7}$	0.017	0.008	0.006	0.005	0.004	
100	$n^{-1/8}$	0.017	0.008	0.006	0.005	0.005	0.053
	$(\log n)^{-1}$	0.016	0.008	0.006	0.005	0.004	
	$n^{-1/7}$	0.022	0.015	0.011	0.009	0.009	
200	$n^{-1/8}$	0.021	0.015	0.011	0.009	0.009	0.071
	$(\log n)^{-1}$	0.021	0.015	0.011	0.009	0.009	
	$n^{-1/7}$	0.010	0.002	0.002	0.003	0.003	
300	$n^{-1/8}$	0.011	0.002	0.002	0.003	0.002	0.064
	$(\log n)^{-1}$	0.010	0.002	0.002	0.003	0.003	
	$n^{-1/7}$	0.013	0.006	0.000	0.000	0.000	
400	$n^{-1/8}$	0.014	0.005	0.000	0.000	0.000	0.068
	$(\log n)^{-1}$	0.014	0.005	0.000	0.000	0.000	
	$n^{-1/7}$	0.009	0.002	0.003	0.000	0.000	
1000	$n^{-1/8}$	0.009	0.002	0.003	0.000	0.000	0.078
	$(\log n)^{-1}$	0.009	0.002	0.003	0.000	0.000	

Next, we compare our test with some existing tests. As before, the null hypothesis is that the true cdf is conave (or, equivalently, the true pdf is nonincreasing), while the alternative is that it is not. We shall consider the D and the P tests in Woodroofe and Sun (1999), the Kolmogorov-Smirnov type test in Durot (2003) and Kulikov and Lopuhaä (2004), labelled as the KS test, and the test based on the R_n statistic with $k = 1$ in Kulikov and Lopuhaä (2004), labelled as the KL test. We design two families of distributions. The first family consists of distributions defined by: for $\lambda \in \mathbf{R}$,

$$F_\lambda(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{e^{\lambda x}-1}{e^{\lambda}-1} & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}, \tag{4.5}$$

where F_λ is understood to be the uniform cdf if $\lambda = 0$. Clearly, F_λ is concave and hence in the null whenever $\lambda \le 0$, but is convex and hence in the alternative if $\lambda > 0$—see Figure 1-(a). Thus, this is a setup well suited for the D test and the P test proposed by Woodroofe and Sun (1999), because the alternative consists of only convex cdfs (or nondecreasing pdfs). As mentioned in Section 3.3, the D and the P tests may perform poorly in detecting distributions that are neither convex nor concave. To verify this point numerically, we consider a family of

TABLE 2
Pointwise Size Control with Non-Strict Concavity Based on (4.2)

		Rescaled					Standard
		t_n					
n	κ_n	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	
	$n^{-1/7}$	0.068	0.051	0.046	0.041	0.036	
100	$n^{-1/8}$	0.069	0.049	0.044	0.041	0.036	0.132
	$(\log n)^{-1}$	0.070	0.052	0.045	0.041	0.036	
	$n^{-1/7}$	0.082	0.061	0.052	0.046	0.045	
200	$n^{-1/8}$	0.084	0.063	0.054	0.045	0.043	0.151
	$(\log n)^{-1}$	0.080	0.063	0.053	0.046	0.044	
	$n^{-1/7}$	0.086	0.066	0.056	0.046	0.040	
300	$n^{-1/8}$	0.085	0.068	0.056	0.047	0.040	0.169
	$(\log n)^{-1}$	0.086	0.067	0.056	0.046	0.040	
	$n^{-1/7}$	0.109	0.081	0.064	0.057	0.053	
400	$n^{-1/8}$	0.106	0.079	0.064	0.057	0.050	0.191
	$(\log n)^{-1}$	0.108	0.079	0.063	0.058	0.051	
	$n^{-1/7}$	0.096	0.071	0.063	0.055	0.046	
1000	$n^{-1/8}$	0.095	0.072	0.063	0.054	0.046	0.205
	$(\log n)^{-1}$	0.093	0.073	0.062	0.056	0.046	

distributions (that belongs to the alternative) defined by: for $\lambda \in (0,1)$,

$$F_\lambda(x) = \begin{cases} 0 & \text{if } x < 0 \\ -(x-\lambda)^2/\lambda + \lambda & \text{if } 0 \le x < \lambda \\ x & \text{if } \lambda \le x < 1 \\ 1 & \text{if } \lambda \ge 1 \end{cases}. \tag{4.6}$$

As shown in Figure 1-(b), F_λ defined in (4.6) is neither concave nor convex on $[0,1]$ for any $\lambda \in (0,1)$, though it is concave on $[0,\lambda]$ and $[\lambda,1]$.

As suggested by Woodroofe and Sun (1999), we choose the penalty parameter c for the D and the P tests to be 0.2, and the corresponding critical values may be found in their Table 1. The critical values for the KS and the KL tests are given in Table 1 in Kulikov and Lopuhaä (2004)—see also Table 1 in Durot (2003). For convenience, if $(t_n, \kappa_n) = (n^{-1/3}, n^{1-/7})$, then we label our test as R37—"R" stands for "the rescaled bootstrap"; if $(t_n, \kappa_n) = (n^{-1/3}, \log^{-1} n)$, then the resulting test is labelled as R3L—"L" stands for "logarithm"; so on and so forth. Below, we try more sample sizes with $n \in \{100, 200, 300, 400, 600, 800, 1000\}$, but only report results for $n \in \{100, 400\}$ in the main text and relegate the rest to Appendix C in Fang (2019).

Tables 5 and 6 record the results for $n \in \{100, 400\}$. Across the choices of (t_n, κ_n), when F_λ is close to being uniform, our test tends to over-reject (under

TABLE 3
Local Size Control with Strict Concavity Based on $f_{1,1/\sqrt{n}}$ from (4.3)

n	κ_n	Rescaled t_n $n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	Standard
	$n^{-1/7}$	0.017	0.009	0.007	0.008	0.006	
100	$n^{-1/8}$	0.016	0.010	0.007	0.007	0.006	0.050
	$(\log n)^{-1}$	0.017	0.008	0.008	0.007	0.006	
	$n^{-1/7}$	0.018	0.006	0.004	0.003	0.003	
200	$n^{-1/8}$	0.018	0.006	0.004	0.003	0.003	0.060
	$(\log n)^{-1}$	0.018	0.006	0.004	0.003	0.003	
	$n^{-1/7}$	0.015	0.009	0.006	0.002	0.002	
300	$n^{-1/8}$	0.014	0.009	0.006	0.002	0.002	0.059
	$(\log n)^{-1}$	0.016	0.009	0.006	0.002	0.002	
	$n^{-1/7}$	0.015	0.002	0.003	0.000	0.000	
400	$n^{-1/8}$	0.015	0.002	0.003	0.000	0.000	0.070
	$(\log n)^{-1}$	0.015	0.002	0.003	0.000	0.000	
	$n^{-1/7}$	0.009	0.002	0.002	0.002	0.002	
1000	$n^{-1/8}$	0.010	0.002	0.002	0.002	0.002	0.070
	$(\log n)^{-1}$	0.011	0.002	0.000	0.002	0.002	

the null) with smaller t_n, especially with $t_n = n^{-1/3}$ and in small samples. All other tests control size well in large samples, but tend to under-reject in small samples. In terms of power, the KS test appears to be the least powerful against local alternatives (i.e., distributions with positive but small λ). The D-test overall appears to be the most powerful one, though the power is comparable to ours and that of the P and the KL tests. For a fair comparison, we also compute the size-adjusted power by subtracting the empirical rejection rates for $\lambda = 0$ (the least favorable case) from the corresponding rejections rates for $\lambda > 0$. The results are recorded in Table 7, from which we see that the power patterns more or less remain. Those results for $n \in \{200, 300, 600, 800, 1000\}$ share similar patterns—see Appendix C in Fang (2019).

Tables 8 and 9 record the rejection rates for $n \in \{100, 400\}$ based on the design (4.6). As expected, the D and the P tests, which are designed to be powerful against convex alternatives, now perform poorly and in most cases have power less than 5%—this is the case even with large sample sizes. The KS and the KL tests appear more powerful than the D and the P tests, but their performance is strictly dominated by our tests across all sample sizes and pairs (t_n, κ_n), with substantial power discrepancies in many cases. We remind the reader that in practice, one can rarely rule out a priori distributions such as those in (4.6). Thus, the simulations suggest that our test is more robust in

TABLE 4
Local Size Control with Non-Strict Concavity Based on $f_{2,1/\sqrt{n}}$ from (4.4)

n	κ_n	Rescaled					Standard
		t_n					
		$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	
100	$n^{-1/7}$	0.070	0.052	0.043	0.041	0.036	
	$n^{-1/8}$	0.070	0.051	0.042	0.038	0.038	0.128
	$(\log n)^{-1}$	0.073	0.051	0.043	0.038	0.034	
200	$n^{-1/7}$	0.112	0.074	0.063	0.051	0.045	
	$n^{-1/8}$	0.111	0.076	0.063	0.049	0.045	0.193
	$(\log n)^{-1}$	0.113	0.075	0.062	0.049	0.042	
300	$n^{-1/7}$	0.098	0.066	0.055	0.050	0.044	
	$n^{-1/8}$	0.099	0.067	0.058	0.049	0.044	0.182
	$(\log n)^{-1}$	0.100	0.070	0.057	0.050	0.044	
400	$n^{-1/7}$	0.113	0.084	0.071	0.061	0.055	
	$n^{-1/8}$	0.113	0.085	0.070	0.062	0.055	0.197
	$(\log n)^{-1}$	0.112	0.083	0.069	0.061	0.056	
1000	$n^{-1/7}$	0.092	0.064	0.055	0.046	0.036	
	$n^{-1/8}$	0.093	0.061	0.055	0.044	0.036	0.196
	$(\log n)^{-1}$	0.094	0.066	0.055	0.045	0.037	

the sense that they are powerful against a larger class of alternatives. Those results for $n \in \{200, 300, 600, 800, 1000\}$ share similar patterns—see Appendix C in Fang (2019).

Admittedly, the simulation results reinforce the importance of the choice of κ_n for our test—the choice of t_n does not appear as important. Overall, they provide comforting numerical evidence that our test is a useful addition to the literature. The associated computation cost is reasonable: with an Intel Xeon CPU E5-1650 v4 3.60 GHz, a single replication based on 1000 samples from the design in (4.5) and 500 bootstrap repetitions is completed in 15 seconds.

5. Conclusion

This paper studies estimation of and inference on a cumulative distribution function with concavity constraint. The estimation results generalize the asymptotic order results in Kiefer and Wolfowitz (1976) to settings with unbounded support and contiguous distributions. These results are not only of interest in their own right, but also useful for conducting inference on the concavity. In particular, in conjunction with the rescaled bootstrap of Dümbgen (1993) and the recent work of Fang and Santos (2019), they allow us to build up a test that controls size, pointwise and locally, even when the distribution function is strictly concave (in
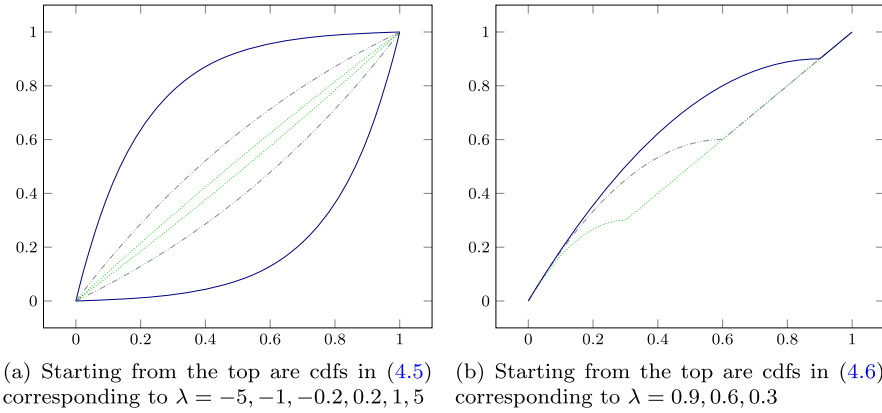
(a) Starting from the top are cdfs in (4.5) corresponding to $\lambda = -5, -1, -0.2, 0.2, 1, 5$

(b) Starting from the top are cdfs in (4.6) corresponding to $\lambda = 0.9, 0.6, 0.3$

Fig 1. *Graphical representations of the cdfs in* (4.5) *and* (4.6)

Table 5
Comparisons with Existing Tests Based on (4.5): $n = 100$

	λ									
	-1	-0.1	-0.05	0	0.05	0.1	0.15	0.2	0.25	1
R37	0.010	0.084	0.106	0.095	0.124	0.132	0.181	0.194	0.235	0.819
R38	0.009	0.085	0.107	0.094	0.123	0.133	0.180	0.187	0.235	0.818
R3L	0.009	0.085	0.107	0.096	0.123	0.135	0.180	0.189	0.236	0.818
R47	0.005	0.062	0.084	0.079	0.103	0.109	0.157	0.165	0.206	0.799
R48	0.004	0.064	0.086	0.080	0.104	0.111	0.156	0.165	0.208	0.799
R4L	0.004	0.061	0.088	0.079	0.102	0.110	0.157	0.165	0.209	0.798
R57	0.003	0.049	0.079	0.076	0.094	0.097	0.149	0.152	0.194	0.789
R58	0.003	0.051	0.078	0.075	0.094	0.099	0.146	0.153	0.195	0.789
R5L	0.003	0.050	0.079	0.075	0.095	0.097	0.145	0.154	0.195	0.789
R67	0.002	0.049	0.075	0.075	0.091	0.091	0.137	0.144	0.187	0.784
R68	0.002	0.049	0.076	0.075	0.091	0.091	0.137	0.145	0.187	0.784
R6L	0.002	0.047	0.075	0.073	0.091	0.090	0.144	0.146	0.188	0.787
R77	0.002	0.047	0.073	0.071	0.087	0.088	0.137	0.143	0.180	0.782
R78	0.003	0.045	0.074	0.071	0.088	0.090	0.139	0.145	0.182	0.783
R7L	0.002	0.045	0.074	0.072	0.089	0.088	0.141	0.143	0.182	0.784
P-Test	0.003	0.032	0.030	0.037	0.064	0.068	0.072	0.079	0.119	0.764
D-Test	0.000	0.018	0.039	0.036	0.061	0.071	0.090	0.108	0.170	0.809
KS-Test	0.000	0.010	0.017	0.018	0.031	0.035	0.048	0.057	0.083	0.642
KL-Test	0.000	0.012	0.020	0.024	0.038	0.042	0.052	0.074	0.109	0.765

TABLE 6
Comparisons with Existing Tests Based on (4.5): $n = 400$

	λ									
	-1	-0.1	-0.05	0	0.05	0.1	0.15	0.2	0.25	1
R37	0.004	0.052	0.077	0.086	0.134	0.189	0.286	0.327	0.420	0.999
R38	0.004	0.053	0.074	0.088	0.134	0.191	0.287	0.326	0.420	0.999
R3L	0.004	0.052	0.075	0.087	0.133	0.191	0.287	0.324	0.422	0.999
R47	0.003	0.033	0.058	0.068	0.107	0.163	0.252	0.296	0.382	0.999
R48	0.003	0.036	0.059	0.065	0.107	0.166	0.252	0.292	0.384	0.999
R4L	0.002	0.035	0.059	0.067	0.107	0.164	0.251	0.296	0.385	0.999
R57	0.003	0.027	0.050	0.061	0.096	0.150	0.237	0.282	0.364	0.999
R58	0.003	0.029	0.053	0.062	0.097	0.150	0.237	0.279	0.365	0.999
R5L	0.003	0.029	0.051	0.061	0.095	0.153	0.237	0.278	0.364	0.999
R67	0.000	0.027	0.048	0.060	0.092	0.144	0.227	0.272	0.349	0.999
R68	0.000	0.024	0.046	0.052	0.080	0.151	0.226	0.276	0.368	0.999
R6L	0.000	0.027	0.050	0.058	0.093	0.146	0.230	0.273	0.354	0.999
R77	0.000	0.023	0.048	0.055	0.087	0.140	0.224	0.268	0.347	0.999
R78	0.000	0.026	0.050	0.053	0.087	0.143	0.223	0.266	0.345	0.999
R7L	0.003	0.024	0.048	0.056	0.088	0.140	0.224	0.265	0.346	0.999
P-Test	0.000	0.024	0.035	0.041	0.078	0.105	0.189	0.234	0.343	0.999
D-Test	0.000	0.013	0.023	0.037	0.065	0.135	0.219	0.294	0.382	0.999
KS-Test	0.000	0.010	0.020	0.025	0.046	0.080	0.145	0.198	0.259	0.999
KL-Test	0.000	0.009	0.018	0.037	0.060	0.102	0.184	0.258	0.340	1.000

which case the test statistic is asymptotically degenerate). Through simulation studies, we find that our test is powerful against a larger class of alternatives, compared to some existing tests such as those in Woodroofe and Sun (1999) that are designed to work against a specific class of alternatives, or those in Durot (2003) and Kulikov and Lopuhaä (2004) that rely on critical values from the least favorable asymptotic distributions. In Appendix B, we show how the testing results may be extended to a general setup that includes regression and hazard rate problems as special cases.

In our limited simulation studies, we also find that the choice of κ_n for the KW selection does not appear as important as the choice of t_n for implementing the rescaled bootstrap. A formal investigation of both choices is, however, beyond the scope of this paper, and therefore left for future study.

Appendix A: Proofs of main results

Proof of Theorem 2.1. Our proof is a mixture of the original one in Kiefer and Wolfowitz (1976) and the "modernized" one given in Balabdaoui and Wellner

TABLE 7
Size-Adjusted Power Comparisons with Existing Tests Based on (4.5)

	λ for $n = 100$						λ for $n = 400$					
	0.05	0.1	0.15	0.2	0.25	1	0.05	0.1	0.15	0.2	0.25	1
R37	0.029	0.037	0.086	0.099	0.140	0.724	0.048	0.103	0.200	0.241	0.334	0.913
R38	0.029	0.039	0.086	0.093	0.141	0.724	0.046	0.103	0.199	0.238	0.332	0.911
R3L	0.027	0.039	0.084	0.093	0.140	0.722	0.046	0.104	0.200	0.237	0.335	0.912
R47	0.024	0.030	0.078	0.086	0.127	0.720	0.039	0.095	0.184	0.228	0.314	0.931
R48	0.024	0.031	0.076	0.085	0.128	0.719	0.042	0.101	0.187	0.227	0.319	0.934
R4L	0.023	0.031	0.078	0.086	0.130	0.719	0.040	0.097	0.184	0.229	0.318	0.932
R57	0.018	0.021	0.073	0.076	0.118	0.713	0.035	0.089	0.176	0.221	0.303	0.938
R58	0.019	0.024	0.071	0.078	0.120	0.714	0.035	0.088	0.175	0.217	0.303	0.937
R5L	0.020	0.022	0.070	0.079	0.120	0.714	0.034	0.092	0.176	0.217	0.303	0.938
R67	0.016	0.016	0.062	0.069	0.112	0.709	0.032	0.084	0.167	0.212	0.289	0.939
R68	0.016	0.016	0.062	0.070	0.112	0.709	0.028	0.099	0.174	0.224	0.316	0.947
R6L	0.018	0.017	0.071	0.073	0.115	0.714	0.035	0.088	0.172	0.215	0.296	0.941
R77	0.016	0.017	0.066	0.072	0.109	0.711	0.032	0.085	0.169	0.213	0.292	0.944
R78	0.017	0.019	0.068	0.074	0.111	0.712	0.034	0.090	0.170	0.213	0.292	0.946
R7L	0.017	0.016	0.069	0.071	0.110	0.712	0.032	0.084	0.168	0.209	0.290	0.943
P-Test	0.027	0.031	0.035	0.042	0.082	0.727	0.037	0.064	0.148	0.193	0.302	0.958
D-Test	0.025	0.035	0.054	0.072	0.134	0.773	0.028	0.098	0.182	0.257	0.345	0.962
KS-Test	0.013	0.017	0.030	0.039	0.065	0.624	0.021	0.055	0.120	0.173	0.234	0.974
KL-Test	0.014	0.018	0.028	0.050	0.085	0.741	0.023	0.065	0.147	0.221	0.303	0.963

(2007) but takes into account that the support of the density function f is potentially unbounded. Following Kiefer and Wolfowitz (1976), we first define interpolating processes for F and $\mathbb{F}_n$. Let $k_n \uparrow \infty$ be a sequence of positive integers (to be chosen). Define $a_j \equiv a_j^{(k_n)} \equiv F^{-1}(j/k_n)$ for $j = 1, \ldots, k_n - 1$, and $a_{k_n} \equiv a_{k_n}^{(k_n)} \equiv F^{-1}(1 - \frac{1}{(2k_n)})$. Moreover, set $a_0 = 0$ and $a_{k_n+1} \equiv a_{k_n+1}^{(k_n)} = \infty$. Let $L^{(k_n)}$ be the piecewise linear on $[a_{j-1}, a_j]$ for $j = 1, \ldots, k_n$ satisfying

$$L^{(k_n)}(a_j^{(k_n)}) = F(a_j^{(k_n)}) \, , \, j = 0, \ldots, k_n \, ,$$

and $L^{(k_n)}(x) = F(x)$ for $x \in [a_{k_n}, a_{k_n+1}]$. Thus, in Figure 2, L^{k_n} is the function that connects $\{a_j, j/k_n\}_{j=0}^{k_n}$ on the interval $[0, a_{k_n}]$ in a piecewise linear way and is identical to $F(x)$ on the tail $[a_{k_n}, a_{k_n+1}]$. Clearly, $L^{(k_n)}$ inherits concavity from F. Using the notation in de Boor (2001, p.31), we may write $L^{(k_n)} = I_2 F$ even though $L^{(k_n)}$ is nonlinear on $[a_{k_n}, a_{k_n+1}]$.

Next, we define $L_n^{(k_n)}$ by: for $x \in [a_j, a_{j+1}]$ and $j = 0, \ldots, k_n$,

$$L_n^{(k_n)}(x) \equiv \mathbb{F}_n(a_j) + \frac{\mathbb{F}_n(a_{j+1}) - \mathbb{F}_n(a_j)}{F(a_{j+1}) - F(a_j)} \{L^{(k_n)}(x) - F(a_j)\} \, .$$

Similarly, write $L_n^{(k_n)} = I_2 \mathbb{F}_n$ and note, for $j = 0, \ldots, k_n + 1$,

$$L_n^{(k_n)}(a_j^{(k_n)}) = \mathbb{F}_n(a_j^{(k_n)}) \, .$$

TABLE 8
Comparisons with Existing Tests Based on (4.6): $n = 100$

	λ								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
R37	0.127	0.221	0.339	0.482	0.652	0.664	0.607	0.386	0.056
R38	0.125	0.223	0.336	0.483	0.656	0.669	0.607	0.387	0.053
R3L	0.127	0.222	0.338	0.483	0.654	0.669	0.608	0.386	0.055
R47	0.102	0.184	0.273	0.406	0.566	0.586	0.503	0.296	0.030
R48	0.103	0.182	0.273	0.407	0.569	0.584	0.499	0.299	0.029
R4L	0.102	0.182	0.275	0.406	0.563	0.582	0.504	0.297	0.031
R57	0.097	0.164	0.255	0.374	0.531	0.537	0.447	0.248	0.024
R58	0.095	0.160	0.252	0.377	0.529	0.537	0.448	0.250	0.024
R5L	0.096	0.162	0.256	0.374	0.527	0.539	0.449	0.247	0.024
R67	0.086	0.153	0.238	0.358	0.494	0.508	0.411	0.213	0.020
R68	0.084	0.152	0.239	0.358	0.497	0.506	0.412	0.212	0.020
R6L	0.086	0.153	0.237	0.357	0.499	0.512	0.412	0.214	0.020
R77	0.087	0.147	0.234	0.346	0.479	0.484	0.389	0.193	0.019
R78	0.085	0.148	0.233	0.344	0.475	0.486	0.389	0.196	0.019
R7L	0.085	0.146	0.234	0.342	0.480	0.483	0.394	0.196	0.019
P-Test	0.015	0.018	0.008	0.009	0.010	0.004	0.003	0.006	0.003
D-Test	0.030	0.048	0.039	0.042	0.032	0.017	0.006	0.004	0.000
KS-Test	0.017	0.055	0.090	0.121	0.154	0.126	0.086	0.021	0.000
KL-Test	0.025	0.050	0.043	0.039	0.015	0.003	0.000	0.000	0.000

Thus, $L_n^{k_n}$ and $\mathbb{F}_n$ intersect at the grid points $\{a_j\}$, just as L^{k_n} and F. Moreover, since $L_n^{(k_n)}$ is an affine transformation of $L^{(k_n)}$ on each of the intervals $[a_j, a_{j+1}]$, the former inherits the piecewise linearity from the latter on $[a_0, a_{k_n}]$. Heuristically, one may thus think of $L_n^{(k_n)}$ as a finite sample analog of $L^{(k_n)}$. We next show that $L_n^{(k_n)}$ also inherits concavity from $L^{(k_n)}$ for a suitable k_n.

<u>STEP 1:</u> For $A_n \equiv \{L_n^{(k_n)}$ is concave on $\mathbf{R}_+\}$ and a suitably chosen sequence $\{k_n\}$, show that

$$\lim_{n\to\infty} P(A_n) = 1 . \tag{A.1}$$

In fact, we shall show that A_n occurs for all n large with probability one. We follow the arguments given by Kiefer and Wolfowitz (1976, p.81-2). For $j = 1, \ldots, k_n$, define the increments:

$$T_{n,j} \equiv \mathbb{F}_n(a_j) - \mathbb{F}_n(a_{j-1}) , \; \Delta a_j \equiv a_j - a_{j-1} .$$

Note that the definition of Δa_j is in line with Balabdaoui and Wellner (2007)

TABLE 9
Comparisons with Existing Tests Based on (4.6): $n = 400$

	λ								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
R37	0.207	0.595	0.938	0.995	1.000	1.000	1.000	1.000	0.718
R38	0.207	0.598	0.935	0.995	1.000	1.000	1.000	1.000	0.721
R3L	0.207	0.596	0.937	0.995	1.000	1.000	1.000	1.000	0.723
R47	0.173	0.516	0.908	0.991	1.000	1.000	1.000	0.995	0.563
R48	0.176	0.517	0.908	0.991	1.000	1.000	1.000	0.995	0.564
R4L	0.175	0.514	0.908	0.991	1.000	1.000	1.000	0.995	0.561
R57	0.157	0.485	0.886	0.987	0.998	1.000	1.000	0.985	0.463
R58	0.155	0.487	0.886	0.989	0.998	0.999	1.000	0.986	0.470
R5L	0.158	0.485	0.887	0.988	0.998	1.000	1.000	0.985	0.467
R67	0.143	0.469	0.876	0.985	0.998	0.999	1.000	0.975	0.394
R68	0.143	0.471	0.875	0.985	0.998	0.999	1.000	0.976	0.393
R6L	0.144	0.468	0.875	0.986	0.998	0.999	1.000	0.977	0.391
R77	0.137	0.462	0.864	0.980	0.997	0.999	1.000	0.966	0.346
R78	0.134	0.461	0.864	0.979	0.997	0.999	1.000	0.967	0.350
R7L	0.136	0.456	0.866	0.981	0.997	0.999	1.000	0.967	0.348
P-Test	0.020	0.014	0.010	0.015	0.010	0.007	0.007	0.005	0.003
D-Test	0.050	0.054	0.042	0.038	0.032	0.012	0.011	0.004	0.000
KS-Test	0.076	0.295	0.689	0.907	0.966	0.977	0.962	0.750	0.056
KL-Test	0.107	0.232	0.356	0.362	0.298	0.083	0.003	0.000	0.000

but differs slightly from Kiefer and Wolfowitz (1976). Since $L_n^{(k_n)}$ is piecewise linear on each $[a_j, a_{j-1}]$ for all $j = 1, \ldots, k_n$, the slope of $L_n^{(k_n)}$ on each such $[a_j, a_{j-1}]$ is precisely $T_{n,j}/\Delta a_j$. In turn, it follows that establishing "asymptotic concavity" amounts to showing a string of inequalities comparing consecutive slopes—we remind the reader that $L_n^{(k_n)}$ is concave on the interval $[a_{k_n}, a_{k_n+1}]$. To formalize the idea, write, for $j = 1, \ldots, k_n - 1$,

$$B_{n,j} \equiv \left\{ \frac{T_{n,j+1}}{\Delta a_{j+1}} \leq \frac{T_{n,j}}{\Delta a_j} \right\} \ , \ B_{n,k_n} \equiv \left\{ \frac{T_{n,k_n+1}}{F(a_{k_n+1}) - F(a_{k_n})} f(a_{k_n}) \leq \frac{T_{n,a_{k_n}}}{\Delta a_{k_n}} \right\} \ ,$$

where $\frac{T_{n,k_n+1}}{F(a_{k_n+1})-F(a_{k_n})} f(a_{k_n})$ is the slope of $L_n^{(k_n)}$ at a_{k_n}. For all $j = 1, \ldots, k_n$, if $B_{n,j}$ holds, then $L_n^{(k_n)}$ stays concave as it moves from $[a_{j-1}, a_j]$ into $[a_j, a_{j+1}]$. Therefore, we obtain the following representation of the event A_n:

$$A_n = \bigcap_{j=1}^{k_n} B_{n,j} \ . \tag{A.2}$$
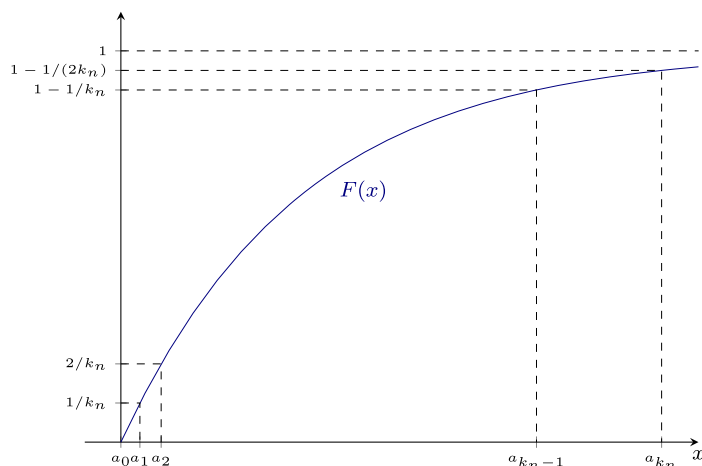
FIG 2. *Construction of the grid points* $a_0, a_1, a_2, \ldots, a_{k_n}$

We next consider the sets $\{B_{n,j}\}$ one by one. The goal is to establish analytically tractable sufficient conditions for each $B_{n,j}$ to occur. Then we shall bound $P(A_n)$ from below by computing $\sum_{j=1}^{k_n} P(B_{n,j}^c)$ (in view of (A.2)), which in turn may be controlled by utilizing those sufficient conditions.

<u>FOR $j = 1, \ldots, k_n - 2$:</u> It is simple to verify that $B_{n,j}$ occurs if

$$|T_{n,i} - \frac{1}{k_n}| \leq \frac{\delta_n}{k_n} \ , \ i = j, j+1 \ , \ \text{and} \ \frac{\Delta a_{j+1}}{\Delta a_j} \geq 1 + 3\delta_n \ , \qquad (A.3)$$

provided $0 \leq \delta_n \leq 1/3$. Indeed, if (A.3) holds, then

$$\frac{T_{n,j+1}}{T_{n,j}} \leq \frac{1/k_n + \delta_n/k_n}{1/k_n - \delta_n/k_n} \leq 1 + 3\delta_n \leq \frac{\Delta a_{j+1}}{\Delta a_j} \ , \qquad (A.4)$$

so that $B_{n,j}$ occurs, where the second inequality holds whenever $0 \leq \delta_n \leq 1/3$. Next, we show that $\frac{\Delta a_{j+1}}{\Delta a_j} \geq 1+3\delta_n$ holds for all n large under an additional rate restriction on δ_n (to be specified)—we shall control the probability of $|T_{n,i} - \frac{1}{k_n}| \leq \frac{\delta_n}{k_n}$ shortly. By Assumption 2.2(i), we may employ Taylor's theorem to conclude that, for some $\xi_{j+1} \in [a_j, a_{j+1}]$,

$$\Delta a_{j+1} = F^{-1}(\frac{j+1}{k_n}) - F^{-1}(\frac{j}{k_n}) = \frac{1}{f(a_j)}\frac{1}{k_n} + \frac{1}{2k_n^2}\frac{-f'(\xi_{j+1})}{f^3(\xi_{j+1})} \ . \qquad (A.5)$$

Moreover, since F^{-1} is strictly convex (because $\frac{d}{dx}F^{-1}(x) = 1/f(F^{-1}(x))$ is strictly increasing) and $\Delta a_j \equiv F^{-1}(\frac{j}{k_n}) - F^{-1}(\frac{j-1}{k_n})$ by definition, the ratio $\Delta a_j/k_n^{-1}$ is no larger than the slope of F^{-1} at the right end point of the interval $[(j-1)/k_n, j/k_n]$ (i.e., j/k_n). Mathematically, this means that

$$\Delta a_j \leq \frac{1}{f(a_j)}\frac{1}{k_n} \ . \qquad (A.6)$$

Combining previous results (A.5) and (A.6) yields

$$\frac{\Delta a_{j+1}}{\Delta a_j} \geq 1 + \frac{1}{2k_n} \frac{-f'(\xi_{j+1})}{f^3(\xi_{j+1})} f(a_j)$$

$$\geq 1 + \frac{1}{2k_n} \frac{-f'(\xi_{j+1})}{f^2(\xi_{j+1})} \geq 1 + \frac{1}{2k_n} \bar{\beta}(\frac{1}{2k_n}) , \quad (A.7)$$

where the second inequality holds since $f(\xi_{j+1}) \leq f(a_j)$ because f is decreasing and $\xi_{j+1} \geq a_j$. By setting $\delta_n \leq \frac{1}{6k_n} \bar{\beta}(\frac{1}{k_n})$, the second part of display (A.3) thus holds. Note that $\bar{\beta}(\epsilon)$ is clearly nondecreasing in ϵ (because the domain of the infimum in (2.1) shrinks as ϵ increases), so $\delta_n \leq 1/3$ for all large n. Moreover, by Assumption 2.2(ii), $\bar{\beta}(\epsilon) > 0$ for all small $\epsilon > 0$ and so $\delta_n > 0$ for large n.

FOR $j = k_n - 1$: By identical arguments in (A.4), we see B_{n,k_n-1} occurs if

$$|T_{n,k_n-1} - \frac{1}{k_n}| \leq \frac{\delta_n}{k_n} , \; |T_{n,k_n} - \frac{1}{2k_n}| \leq \frac{\delta_n}{2k_n} ,$$

$$\text{and } 2\frac{\Delta a_{k_n}}{\Delta a_{k_n-1}} \geq 1 + 3\delta_n , \quad (A.8)$$

provided $0 \leq \delta_n \leq 1/3$. Indeed, the first two inequalities in (A.8) imply

$$\frac{T_{n,k_n}}{T_{n,k_n-1}} \leq \frac{1/(2k_n) + \delta_n/(2k_n)}{1/k_n - \delta_n/k_n} = \frac{1}{2} \cdot \frac{1+\delta_n}{1-\delta_n} \leq \frac{1+3\delta_n}{2} \leq \frac{\Delta a_{k_n}}{\Delta a_{k_n-1}} ,$$

as claimed, where the second inequality holds provided $0 \leq \delta_n \leq 1/3$. By arguments analogous to those leading to (A.7), the last displayed part of (A.8) holds if $\delta_n \leq \frac{1}{12k_n} \bar{\beta}(\frac{1}{2k_n}) \leq 1/3$. Indeed, by Taylor's theorem (as in (A.5)) we have, for some $\xi_{k_n} \in [a_{k_n-1}, a_{k_n}]$,

$$\Delta a_{k_n} = F^{-1}(1 - \frac{1}{2k_n}) - F^{-1}(1 - \frac{1}{k_n})$$

$$= \frac{1}{f(a_{k_n-1})} \frac{1}{2k_n} + \frac{1}{8k_n^2} \frac{-f'(\xi_{k_n})}{f^3(\xi_{k_n})} , \quad (A.9)$$

Moreover, (A.6) holds for $j = k_n - 1$ as well, i.e.,

$$\Delta a_{k_n-1} \leq \frac{1}{f(a_{k_n-1})} \frac{1}{k_n} . \quad (A.10)$$

A combination of (A.9) and (A.10), together with the definition of $\bar{\beta}$, f being decreasing, then implies that

$$\frac{\Delta a_{k_n}}{\Delta a_{k_n-1}} \geq \frac{1}{2} + \frac{1}{8k_n} \frac{-f'(\xi_{k_n})}{f^3(\xi_{k_n})} f(a_{k_n-1})$$

$$\geq \frac{1}{2} + \frac{1}{8k_n} \bar{\beta}(\frac{1}{2k_n}) \geq \frac{1}{2}(1 + 3\delta_n) , \quad (A.11)$$

where the last inequality holds if $\delta_n \leq \frac{1}{12k_n}\bar{\beta}(\frac{1}{2k_n})$. We shall control the probabilities of the first two events in (A.8) shortly.

FOR $j = k_n$: This case is easier because B_{n,k_n} occurs whenever

$$|T_{n,k_n} - \frac{1}{2k_n}| \leq \frac{\delta_n}{2k_n} \ , \ T_{n,k_n+1} \leq \frac{1-\delta_n}{2k_n} \ . \tag{A.12}$$

To see this, note that, as in result (A.6), convexity of F^{-1} and $\Delta a_{k_n} \equiv F^{-1}(1 - 1/(2k_n)) - F^{-1}(1 - 1/k_n)$ together imply that the ratio $\Delta a_j/(1/(2k_n))$ is no larger than the slope of F^{-1} at $1 - 1/(2k_n)$, i.e.,

$$\Delta a_{k_n} \leq \frac{1}{f(a_{k_n})} \frac{1}{2k_n} \ .$$

Hence, by construction of the grid points, we have

$$\frac{F(a_{k_n+1}) - F(a_{k_n})}{\Delta a_{k_n}} \frac{1}{f(a_{k_n})} = \frac{1/(2k_n)}{\Delta a_{k_n} f(a_{k_n})} \geq 1 \ . \tag{A.13}$$

It follows from (A.13) that, whenever (A.12) occurs,

$$\frac{T_{n,k_n}}{\Delta a_{k_n}} \frac{F(a_{k_n+1}) - F(a_{k_n})}{f(a_{k_n})} \geq \frac{1-\delta_n}{2k_n} \geq T_{n,k_n+1} \ .$$

At this point, we note that δ_n should be selected to satisfy (i) $0 < \delta_n \leq 1/3$, (ii) $\delta_n \leq \bar{\beta}(1/k_n)/(6k_n)$, and (iii) $\delta_n \leq \bar{\beta}(1/(2k_n))/(12k_n)$. Since $\bar{\beta}(\epsilon)$ decreases as $\epsilon \downarrow 0$, we may thus set $\delta_n = \bar{\beta}(1/(2k_n))/(12k_n)$ in what follows, so that (i)(ii)(iii) hold simultaneously for all n large. Note that $\delta_n \downarrow 0$ since $k_n \uparrow \infty$ as required (and to be constructed). We may now obtain from (A.2), and results (A.3), (A.8) and (A.12) that, for all n large enough so that $\delta_n \leq 1/3$,

$$P(A_n^c) \leq \sum_{j=1}^{k_n} P(B_{n,j}^c) \leq \sum_{j=1}^{k_n-1} 2P(|T_{n,j} - \frac{1}{k_n}| > \frac{\delta_n}{k_n})$$

$$+ 2P(|T_{n,k_n} - \frac{1}{2k_n}| > \frac{\delta_n}{2k_n}) + P(T_{n,k_n+1} \leq \frac{1-\delta_n}{2k_n}) \ . \tag{A.14}$$

Following Balabdaoui and Wellner (2007), we may apply Theorem 1.1.2 in Shorack and Wellner (1986) and Lemma 5.2 in Balabdaoui and Wellner (2007) to conclude that, for $p_n \equiv k_n^{-1}$ and each $j = 1, \ldots, k_n - 1$,

$$P(|T_{n,j} - \frac{1}{k_n}| > \frac{\delta_n}{k_n}) \leq 2\exp\{-\frac{1}{2}np_n\delta_n^2(1 + o(1)) \ , \tag{A.15}$$

where the $o(1)$ term depends only on $\{\delta_n\}$ (and nothing else, including j). The same arguments also imply

$$P(|T_{n,k_n} - \frac{1}{2k_n})| > \frac{\delta_n}{2k_n}) \leq 2\exp\{-\frac{1}{4}np_n\delta_n^2(1 + o(1)) \ , \tag{A.16}$$

where the $o(1)$ term depends only on δ_n. For the last term in (A.14), we need somewhat different arguments, which is important in handling the unbounded support. Specifically, by Theorem 1.1.2 in Shorack and Wellner (1986),

$$P(T_{n,k_n+1} \leq \frac{1-\delta_n}{2k_n}) = P(\frac{1/(2k_n)}{\mathbb{G}_n(1/(2k_n))} \geq \frac{1}{1-\delta_n}) \,, \qquad (A.17)$$

where $\mathbb{G}_n$ is the empirical cdf defined by n i.i.d. uniform random variables. Inequality 10.3.2-(6) in Shorack and Wellner (1986, p.415-6) then implies

$$P(\frac{1/(2k_n)}{\mathbb{G}_n(1/(2k_n))} \geq \frac{1}{1-\delta_n}) \leq P(\sup_{t=1/(2k_n)}^{1} \frac{t}{\mathbb{G}_n(t)} \geq \frac{1}{1-\delta_n})$$

$$\leq \exp\{-\frac{1}{2}np_nh(1-\delta_n)\} \,, \quad (A.18)$$

where $h(x) \equiv x(\log x - 1) + 1$ for $x > 0$. By L'Hôpital's rule, $h(1+x) = \frac{1}{2}x^2(1+o(1))$ as $x \to 0$. Since $\delta_n \downarrow 0$ by our choice, we may thus conclude from results (A.17) and (A.18) that

$$P(T_{n,k_n+1} \leq \frac{1-\delta_n}{2k_n}) \leq \exp\{-\frac{1}{4}np_n\delta_n^2(1+o(1))\} \,, \qquad (A.19)$$

where the $o(1)$ term again depends only on δ_n. Finally, we conclude from results (A.14), (A.15), (A.16) and (A.19), together with the choice of δ_n, that

$$P(A_n^c) \leq 5k_n \exp\{-\frac{1}{4}np_n\delta_n^2(1+o(1))$$

$$\leq 5k_n \exp\{-\frac{1}{720}nk_n^{-3}\bar{\beta}^2(\frac{1}{2k_n}))\} \,, \quad (A.20)$$

for all n large so that the $1+o(1)$ term is larger than $4/5$.

Now, we may choose k_n to be the integer part of ν that satisfies

$$\frac{\nu^3}{\bar{\beta}^2(1/(2\nu))} = \frac{3}{7}\frac{1}{720}\frac{n}{\log n} \,, \qquad (A.21)$$

where $\nu \geq 1$. To see that a solution to (A.21) exists, we note several facts. First, $\epsilon \mapsto \bar{\beta}(\epsilon)$ being nondecreasing implies that $\nu \mapsto \nu^3/\bar{\beta}^2(1/(2\nu))$ is strictly increasing and diverges to infinity as $\nu \uparrow \infty$. Second, $\epsilon \mapsto \bar{\beta}(\epsilon)$ in (2.1) is continuous and hence so is $\nu \mapsto \nu^3/\bar{\beta}^2(1/(2\nu))$. Indeed, Assumption 2.2(i) and $f(x) > 0$ for each $x \in \mathbf{R}_+$ (by Assumption 2.1(ii)) implies that $x \mapsto -f'(x)/f^2(x)$ is continuous on $\mathbf{R}_+$. Moreover, the correspondence $\varphi : (0,1) \twoheadrightarrow \mathbf{R}_+$ defined by $\varphi(\epsilon) = [0, F^{-1}(1-\epsilon)]$ is compact-valued, and continuous (justified by first writing φ as the composition of the singleton-valued continuous map $\epsilon \mapsto F^{-1}(1-\epsilon)$ and the continuous correspondence $x \twoheadrightarrow [0,x]$, and then invoking Theorem 17.23 in Aliprantis and Border (2006)). Berge's maximum theorem—see, for example, Theorem 17.31 in Aliprantis and Border (2006)—then implies that $\bar{\beta}(\epsilon)$ is continuous in ϵ. Third, the right hand side of (A.21) diverges to infinity as $n \to \infty$.

Thus, for each fixed n that is sufficiently large, the right hand side of (A.21) falls within the range $\{\nu^3/\bar{\beta}^2(1/(2\nu)) : \nu \geq 1\}$.

The above three facts together imply that there is a unique $\nu = \nu_n$ that satisfies equation (A.21) for all n large and diverges to infinity as $n \to \infty$. To evaluate the order of ν_n, we note that, by Assumption 2.2(ii),

$$\nu_n^3 = \frac{3}{7}\frac{1}{720}\frac{n}{\log n}\bar{\beta}^2(\frac{1}{2\nu_n}) \asymp \frac{n}{\log n}\nu_n^{-2\tau} \ , \tag{A.22}$$

implying that $\nu_n \asymp (n/\log n)^{1/(2\tau+3)}$. Since k_n is the integer part of ν_n, we have $k_n \asymp (n/\log n)^{1/(2\tau+3)}$, which diverges to infinity as $n \uparrow \infty$ since $2\tau + 3 > 0$ by Assumption 2.2(iii). Since the bound in (A.20) is weakly increasing in k_n, we may thus conclude by results (A.20) and (A.22) that, for all n large,

$$P(A_n^c) \leq 5\nu_n \exp\{-\frac{1}{720}n\nu_n^{-3}\bar{\beta}^2(\frac{1}{2\nu_n}))$$
$$= 5\nu_n \exp\{-\frac{7}{3}\log n\} \lesssim n^{1/3}n^{-7/3} = n^{-2} \to 0 \ . \tag{A.23}$$

Thus, $P(A_n) \to 1$. Since $\sum_{i=1}^n P(A_n^c) = \sum_{i=1}^n n^{-2} < \infty$, the first Borel-Cantelli lemma implies that, with probability one, A_n is concave for all n large.

<u>STEP 2</u>: Relate $\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty$ to the interpolating process $L^{(k_n)}$ and $L_n^{(k_n)}$. For notational simplicity, define

$$B_n \equiv \sup_{x \in [a_{k_n}, a_{k_n+1}]} |\mathbb{F}_n(x) - L_n^{(k_n)}(x)| \ ,$$
$$D_n \equiv \sup_{x \in [0, a_{k_n}]} |\mathbb{F}_n(x) - I_2\mathbb{F}_n(x) - \{F(x) - I_2F(x)\}| \ ,$$
$$E_n \equiv \|F - L^{(k_n)}\|_\infty \ .$$

By the triangle inequality and Lemma 2.2 in Durot and Tocquet (2003), we have

$$\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty \leq \|\hat{\mathbb{F}}_n - L_n^{(k_n)} - \{\mathbb{F}_n - L_n^{(k_n)}\}\|_\infty$$
$$\leq \|\hat{\mathbb{F}}_n - L_n^{(k_n)}\|_\infty + \|\mathbb{F}_n - L_n^{(k_n)}\|_\infty \leq 2\|\mathbb{F}_n - L_n^{(k_n)}\|_\infty \ . \tag{A.24}$$

Invoking the triangle inequality once again we in turn have from (A.24) that

$$\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty \leq 2\|\mathbb{F}_n - L_n^{(k_n)} - \{F - L^{(k_n)}\}\|_\infty + 2\|F - L^{(k_n)}\|_\infty$$
$$\leq 2(B_n + D_n + E_n) \ , \tag{A.25}$$

where the second inequality exploited $F(x) = L^{(k_n)}(x)$ for all $x \in [a_{k_n}, a_{k_n+1}]$.

<u>STEP 3</u>: Conclude by controlling the orders of B_n, D_n and E_n separately. To this end, we introduce the concept of modulus of continuity following de Boor (2001, p.25). For a generic function $g : [a, b] \to \mathbf{R}$, let

$$\omega(g; h) \equiv \sup\{|g(x) - g(y)| : x, y \in [a, b], |x - y| \leq h\} \ .$$

The treatment of D_n follows closely Balabdaoui and Wellner (2007), which we include here for completeness. Note that

$$D_n \leq \omega(\mathbb{F}_n - F; |a|) \stackrel{d}{=} n^{-1/2}\omega(\mathbb{U}_n; p_n) , \qquad (A.26)$$

where the inequality is due to result (18) in de Boor (2001, p.36) with $|a| \equiv \max_{j=1}^{k}(a_j - a_{j-1})$, and $\stackrel{d}{=}$ denotes equality in distribution and is due to Theorem 1.1.2 in Shorack and Wellner (1986), with $\mathbb{U}_n \equiv \sqrt{n}\{\mathbb{G}_n - I\}$ the empirical process of n i.i.d. Uniform$(0,1)$ random variables as defined in Shorack and Wellner (1986). By Theorem 14.2.1 in Shorack and Wellner (1986), we have

$$\lim_{n\to\infty} \frac{\omega(\mathbb{U}_n; p_n)}{\sqrt{2p_n \log(1/p_n)}} = 1 \text{ a.s.}, \qquad (A.27)$$

provided $p_n \to 0$, $np_n \to \infty$, $\log(1/p_n)/\log(\log n) \to \infty$, and $\log(1/p_n)/(np_n) \to 0$. To see that these rate conditions on $p_n \equiv k_n^{-1}$ are indeed met, we first note that $p_n \asymp ((\log n)/n)^{1/(2\tau+3)}$. Thus, trivially, $p_n \to 0$ and

$$np_n = n^{\frac{2\tau+2}{2\tau+3}}(\log n)^{\frac{1}{2\tau+3}} \to \infty , \qquad (A.28)$$

as $n \to \infty$, since $\frac{2\tau+2}{2\tau+3} > 0$ by Assumption 2.2(ii). The third rate condition is also trivial because $\log(1/p_n)$ is of order $\log(n) - \log(\log n)$, diverging to infinity faster than $\log(\log n)$. For the fourth condition, $\log(1/p_n)$ grows at a logarithmic rate while np_n grows at a polynomial rate—note $\tau + 1 > 0$ and so $\frac{2\tau+2}{2\tau+3} > 0$. Now, by simple algebra, we may conclude from result (A.27) that

$$D_n = O_{a.s.}(n^{-1/2}\sqrt{p_n \log(1/p_n)}) = O_{a.s.}(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}) . \qquad (A.29)$$

We now turn to B_n which is important in taking care of the unbounded support. Heuristically, B_n measures the interpolating error for $\mathbb{F}_n$ on the (unbounded) right tail $[a_{k_n}, a_{k_n+1}]$ of the support.

Plugging the expression of $L_n^{(k_n)}$ into B_n we have,

$$\begin{aligned} B_n &= \sup_{x\in[a_{k_n},a_{k_n+1}]} |\mathbb{F}_n(a_{k_n}) + \frac{\mathbb{F}_n(a_{k_n+1}) - \mathbb{F}_n(a_{k_n})}{F(a_{k_n+1}) - F(a_{k_n})}\{L^{(k_n)}(x) - F(a_{k_n})\} - \mathbb{F}_n(x)| \\ &\leq \sup_{x\in[a_{k_n},a_{k_n+1}]} |\mathbb{F}_n(x) - \mathbb{F}_n(a_{k_n})| + |\mathbb{F}_n(a_{k_n+1}) - \mathbb{F}_n(a_{k_n})| \\ &= 2\{\mathbb{F}_n(a_{k_n+1}) - \mathbb{F}_n(a_{k_n})\} , \end{aligned} \qquad (A.30)$$

where the inequality follows by $L^{(k_n)} = F$ on $[a_{k_n}, a_{k_n+1}]$ so that $|L^{(k_n)}(x) - F(a_{k_n})|/|F(a_{k_n+1}) - F(a_{k_n})| \leq 1$, and the second equality follows by monotonicity of $\mathbb{F}_n$. Recall that $T_{n,k_n+1} \equiv \mathbb{F}_n(a_{k_n+1}) - \mathbb{F}_n(a_{k_n})$, which, based on our previous results, satisfies: for all n large,

$$P(T_{n,k_n+1} \leq \frac{1-\delta_n}{2k_n}) \lesssim n^{1/3}n^{-7/3} = n^{-2} . \qquad (A.31)$$

By $\delta_n \downarrow 0$ and the first Borel-Cantelli lemma, we obtain from (A.31) that

$$T_{n,k_n+1} = O_{a.s.}(k_n^{-2}) = O_{a.s.}((\log n)/n)^{2/(2\tau+3)}) , \qquad (\text{A.32})$$

where the second equality follows by $k_n \asymp (n/\log n)^{1/(2\tau+3)}$. It follows from results (A.30) and (A.32) that

$$B_n = O_{a.s.}((\log n)/n)^{2/(2\tau+3)}) . \qquad (\text{A.33})$$

Finally, consider E_n which controls the interpolation error for F using $L^{(k_n)}$. Recall that $L^{(k_n)}$ is piecewise linear on $[0, a_{k_n}]$, and equal to F on $[a_{k_n}, a_{k_n+1}]$. It follows by result (5) in de Boor (2001, p.32), Assumption 2.2, and Theorem 1 in Burchard (1974) (see also Theorem 1 in de Boor (1973)) that

$$E_n = \|F - L^{(k_n)}\|_\infty \leq Ck_n^{-2} \left[\int_0^{a_{k_n}} |f'(x)|^{1/2} \, \mathrm{d}x \right]^2$$
$$\leq Ck_n^{-2}\|f'\|_{1/2} = O((\log n)/n)^{2/(2\tau+3)}) , \qquad (\text{A.34})$$

where C is a constant. If $\tau \geq 0$, then $\frac{2}{2\tau+3} \leq \frac{\tau+2}{2\tau+3}$; if $-1 < \tau < 0$, then $\frac{2}{2\tau+3} > \frac{\tau+2}{2\tau+3}$. Therefore, combining results (A.29), (A.33), and (A.34), together with STEP 1, we may establish the theorem. □

Proof of Theorem 2.2. Let $\{k_n\}$ be the same sequence that is determined by (A.22). Define the grid points $\{a_j\}_{j=0}^{k_n+1}$ and the interpolation processes $L^{(k_n)}$ and $L_n^{(k_n)}$ in the same fashion as before (based on F). The proof analogously consists of three steps as in the proof of Theorem 2.1. One of the key observations is that Theorem 1.1.2 in Shorack and Wellner (1986) holds for any i.i.d. $\{X_i\}_{i=1}^n$ whose common distribution function is allowed to depend on n.

STEP 1: For $A_n \equiv \{L_n^{(k_n)} \text{ is concave on } \mathbf{R}_+\}$ (as before), show that (A.1) holds. This is accomplished by the same arguments as before. In particular, the arguments preceding (A.20) are purely algebraic, Theorem 1.1.2 in Shorack and Wellner (1986) (for result (A.20)) holds for any n, and Lemma 5.2 in Balabdaoui and Wellner (2007) and Inequality 10.3.2 in Shorack and Wellner (1986) are concerned with the uniform empirical process whose relations to $\mathbb{F}_n$ only depend on the i.i.d. assumption (for each n) as characterized by Theorem 1.1.2 in Shorack and Wellner (1986). This last observation means that, again, the common cdf shared by $X_1, \ldots, X_n$ can depend on n, as in Assumption 2.3.

STEP 2: For B_n, D_n and E_n defined as before, show that (A.25) holds. This is a consequence of the triangle inequality and the fact $F(x) = L^{(k_n)}(x)$ for all $x \in [a_{k_n}, a_{k_n+1}]$. The arguments are thus the same as before.

STEP 3: Control B_n, D_n and E_n. The treatment of B_n is the same as before, because Theorem 1.1.2 in Shorack and Wellner (1986) is still applicable which then allows one to invoke Inequality 10.3.2 in Shorack and Wellner (1986). The

treatment of E_n is merely a consequence of the same approximation result, namely, Theorem 1 in Burchard (1974). It thus remains to consider D_n.

Let $F_n \equiv F(P_{1/\sqrt{n}})$ (the true underlying cdf). We then have:

$$
\begin{aligned}
D_n &\equiv \sup_{x \in [0, a_{k_n}]} |\mathbb{F}_n(x) - I_2 \mathbb{F}_n(x) - \{F(x) - I_2 F(x)\}| \\
&\leq \omega(\mathbb{F}_n - F; |a|) \leq \omega(\mathbb{F}_n - F_n; |a|) + \omega(F_n - F; |a|) \\
&\overset{d}{=} n^{-1/2} \omega(\mathbb{U}_n; p_n) + \omega(F_n - F; |a|) ,
\end{aligned}
\tag{A.35}
$$

where the suprema in the ω's above are all taken over $[0, a_{k_n}]$, $|a| \equiv \max_{j=1}^{k}(a_j - a_{j-1})$, the second inequality follows by the subadditivity of $g \mapsto \omega(g, |a|)$ (as an implication of the supremum operator), and $\overset{d}{=}$ is due to Theorem 1.1.2 in Shorack and Wellner (1986). Since $n^{-1/2}\omega(\mathbb{U}_n; p_n) = O_{a.s.}(((\log n)/n)^{\frac{\tau+2}{2\tau+3}})$ as before (see (A.29)), it remains to analyze $\omega(F_n - F; |a|)$.

Let $x_n, y_n \in [0, a_{k_n}]$ with $|x_n - y_n| \leq a$ and $|F(x_n) - F(y_n)| \leq p_n$. Define $f_n \equiv \mathrm{d}P_{1/\sqrt{n}}/\mathrm{d}\mu$ and $f \equiv dP/d\mu$ with μ the Lebesgue measure, which exist by the definition of $\mathbf{P}$. Since $\{P_{1/\sqrt{n}}\} \subset \mathbf{P} \subset \mathbf{M}$, we then obtain by the fundamental theorem of calculus and Jensen's inequality that:

$$
\begin{aligned}
|\{F_n(x_n) - F(x_n)\} - \{F_n(y_n) - F(y_n)\}| &\leq \int_{\mathbf{R}_+} 1\{[x_n, y_n]\}|f_n - f|\,\mathrm{d}\mu \\
&= \int_{\mathbf{R}_+} 1\{[x_n, y_n]\}|f_n - f|1\{f = 0\}\,\mathrm{d}\mu + \int_{\mathbf{R}_+} 1\{[x_n, y_n]\}|f_n - f|1\{f \neq 0\}\,\mathrm{d}\mu \\
&\leq \int_{\mathbf{R}_+} f_n 1\{f = 0\}\,\mathrm{d}\mu + \int_{f \neq 0} 1\{[x_n, y_n]\}|\frac{f_n}{f} - 1|\,\mathrm{d}P ,
\end{aligned}
\tag{A.36}
$$

where the equality follows by the linearity of integrals, and the second inequality exploited $f \cdot 1\{f = 0\} = 0$ in the first integral. Here, $1\{f = 0\}$ and $1\{f \neq 0\}$ are the indicator functions of the sets $\{f = 0\}$ and $\{f \neq 0\}$ respectively. By Proposition A.5.3(E) in Bickel et al. (1998, p.459), we have

$$
\int_{\mathbf{R}_+} f_n 1\{f = 0\}\,\mathrm{d}\mu = o(n^{-1}) = O(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}) .
\tag{A.37}
$$

For each $n \in \mathbf{N}$, let $r_n : \mathbf{R} \to \mathbf{R}$ be defined by (on $\{f \neq 0\}$):

$$
\frac{f_n}{f} = 1 + n^{-1/2}h + n^{-1/2}r_n .
$$

By the triangle inequality and the definition of r_n, we then have

$$
\begin{aligned}
&\int_{f \neq 0} 1\{[x_n, y_n]\}|\frac{f_n}{f} - 1|\,\mathrm{d}P \\
&\leq \int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n|\,\mathrm{d}P + \int_{\mathbf{R}_+} 1\{[x_n, y_n]\}|n^{-1/2}h|\,\mathrm{d}P
\end{aligned}
$$

$$\leq \int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n| \, \mathrm{d}P + n^{-1/2}\{P([x_n, y_n])\}^{1/2}\{\int |h|^2 \, \mathrm{d}P\}^{1/2}$$

$$\leq \int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n| \, \mathrm{d}P + n^{-1/2}p_n^{1/2}\{\int h^2 \, \mathrm{d}P\}^{1/2}$$

$$= \int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n| \, \mathrm{d}P + O(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}) , \qquad (A.38)$$

where the second inequality follows by the Cauchy-Schwarz inequality, the third inequality by $P([x_n, y_n]) = F(y_n) - F(x_n) \leq p_n$, and the last step by $p_n \asymp ((\log n)/n)^{1/(2\tau+3)}$ and $Ph^2 < \infty$. Next, note

$$\int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n| \, \mathrm{d}P$$

$$\leq \int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n|1\{|n^{-1/2}r_n| > 1\} \, \mathrm{d}P$$

$$+ \int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n|1\{|n^{-1/2}r_n| \leq 1\} \, \mathrm{d}P . \quad (A.39)$$

For the first term on the right side above, we have

$$\int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n|1\{|n^{-1/2}r_n| > 1\} \, \mathrm{d}P$$

$$\leq \int_{f \neq 0} |n^{-1/2}r_n|1\{|n^{-1/2}r_n| > 1\} \, \mathrm{d}P$$

$$= o(n^{-1}) = O(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}) , \qquad (A.40)$$

where the first equality follows from Pfanzagl (1985, p.27, lines 5-8), the equivalence between pathwise/Hellinger differentiability and a notion of weak differentiability (which implies the condition (1.2.11') on p.25 in Pfanzagl (1985)). For the second term, by the Cauchy-Schwarz inequality, we have

$$\int_{f \neq 0} 1\{[x_n, y_n]\}|n^{-1/2}r_n|1\{|n^{-1/2}r_n| \leq 1\} \, \mathrm{d}P$$

$$\leq p_n^{1/2}\{\int_{f \neq 0} |n^{-1/2}r_n|^2 1\{|n^{-1/2}r_n| \leq 1\} \, \mathrm{d}P\}^{1/2}$$

$$= p_n^{1/2}n^{-1/2}\{\int_{f \neq 0} |r_n|^2 1\{|n^{-1/2}r_n| \leq 1\} \, \mathrm{d}P\}^{1/2}$$

$$= p_n^{1/2}n^{-1/2}o(1) = O(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}) , \qquad (A.41)$$

where the second equality again follows from Pfanzagl (1985, p.27) but now using the condition (1.2.11") on p.25 in Pfanzagl (1985). Combining results (A.36), (A.37), (A.38), (A.39), (A.40), and (A.41), we finally obtain that

$$|\{F_n(x_n) - F(x_n)\} - \{F_n(y_n) - F(y_n)\}| = O(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}) , \qquad (A.42)$$

and hence, in view of (A.35),

$$D_n = O_{a.s.}(((\log n)/n)^{\frac{\tau+2}{2\tau+3}}) . \qquad (A.43)$$

This completes the proof of the theorem. □

Proof of Lemma 3.1. By Assumption 2.1(i), we obtain by Example 19.6 in van der Vaart (1998) that $\sqrt{n}\{\mathbb{F}_n - F\} \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathbf{R}_+)$, where $\mathbb{G} \equiv \mathbb{B} \circ F$—see also van der Vaart (1998, p.266) for a brief discussion of the limit $\mathbb{G}$. Define a map $\psi : \ell^\infty(\mathbf{R}_+) \to \ell^\infty(\mathbf{R}_+)$ by $\psi(g) = \mathcal{M}g - g$ for any $g \in \ell^\infty(\mathbf{R}_+)$. Then by definition $\phi(g) = \|\psi(g)\|_p$. Since F is concave under H_0, Proposition 2.1 in Beare and Fang (2017) implies that ψ is Hadamard directionally differentiable tangentially to $C_0(\mathbf{R}_+)$ with the derivative $\psi'_F : C_0(\mathbf{R}_+) \to \mathbf{R}$ given by $\phi'_F(h) = \mathcal{M}'_F h - h$, for all $h \in C_0(\mathbf{R}_+)$. Moreover, $\|\cdot\|_p$ is Hadamard directionally differentiable at 0 since $(\|0 + t_n h_n\|_p - \|0\|_p)/t_n = \|h_n\|_p \to \|h\|_p$ whenever $t_n \downarrow 0$ and $h_n \to h$ in $\ell^\infty(\mathbf{R}_+)$. By the chain rule as in Proposition 3.6 in Shapiro (1990), we may therefore conclude that $\phi : \ell^\infty(\mathbf{R}_+) \to \mathbf{R}$ is Hadamard directionally differentiable tangentially to $C_0(\mathbf{R}_+)$ with the derivative $\phi'_F : C_0(\mathbf{R}_+) \to \mathbf{R}$ given by $\phi'_F(h) = \|\mathcal{M}'_F h - h\|_p$. This, in view of Theorem 2.1 in Fang and Santos (2019), implies that: under H_0 (and so $\phi(F) = 0$),

$$\sqrt{n}\phi(\mathbb{F}_n) = \sqrt{n}\{\phi(\mathbb{F}_n) - \phi(F)\} \xrightarrow{L} \phi'_F(\mathbb{G}) = \|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p . \qquad (A.44)$$

This completes the proof of the lemma. □

Proof of Theorem 3.1. Consider first the case when F is strictly concave on $\mathbf{R}_+$. By Assumption 3.1(ii), $n^{-2/3}(\log n)^{2/3}/(n^{-1/2}\kappa_n) \to 0$ as $n \to \infty$. In turn, we may thus have by Theorem 2.1 that

$$\limsup_{n \to \infty} P(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{1-\alpha}) \leq \limsup_{n \to \infty} P(\sqrt{n}\phi(\mathbb{F}_n) > \kappa_n)$$

$$= \limsup_{n \to \infty} P(\frac{\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_\infty}{n^{-1/2}\kappa_n} > 1) = 0 . \qquad (A.45)$$

Now suppose that F is non-strictly concave. In this case, consistency of the rescaled bootstrap is justified by Proposition 2 in Dümbgen (1993). Alternatively, the bootstrap consistency follows by Lemma S.3.8 (for consistency of $\hat{\mathcal{M}}'_n$ as an derivative estimator of $\mathcal{M}'_F$) and Theorem 3.2 in Fang and Santos (2019). In any case, $\hat{c}^*_{1-\alpha} \xrightarrow{p} c^*_{1-\alpha}$ by Assumption 3.2 and Corollary 3.2 in Fang and Santos (2015). Therefore, it follows that

$$\limsup_{n \to \infty} P(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{1-\alpha}) \leq \limsup_{n \to \infty} P(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}^*_{1-\alpha})$$

$$\leq P(\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p \geq c_{1-\alpha}) = \alpha , \qquad (A.46)$$

where the second inequality follows from Slutsky's lemma (so that $\sqrt{n}\phi(\mathbb{F}_n) - \hat{c}^*_{1-\alpha} \xrightarrow{L} \|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p - c_{1-\alpha}$) and the portmanteau theorem, and the last step is due to Assumption 3.2. This proves the first claim of the theorem.

We now turn to the second part of the theorem and suppose that F is not concave. First, we show that $\hat{c}^*_{1-\alpha} = O_p(1)$ regardless of whether F is concave or not, where $O_p(1)$ means "bounded in probability." To this end, note that by Lemma 2.2 in Durot and Tocquet (2003), for any $h \in \ell^\infty(\mathbf{R}_+)$,

$$\|\hat{\mathcal{M}}'_n(h))\|_p \le \|\frac{\mathcal{M}(\mathbb{F}_n + t_n h) - \mathcal{M}(\mathbb{F}_n)}{t_n}\|_p \lesssim \|h\|_\infty , \qquad (A.47)$$

where $\lesssim$ means "smaller than or equal to up to a universal constant". Here, the constant only depends on the (known) weighting function g if $p \in [1, \infty)$. In turn, result (A.47) implies that

$$\begin{aligned}
\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) &- \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p \\
&\le \|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})\|_p + \|\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p \\
&\lesssim 2\|\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_\infty .
\end{aligned} \qquad (A.48)$$

Moreover, Theorem 3.1 in Giné and Zinn (1990) and Proposition 10.7 in Kosorok (2008) yields $\|\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_\infty = O_p(1)$ outer almost surely. This, together with result (A.48) and Lemma 3 in Cheng and Huang (2010), implies that:

$$\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p = O_p(1) , \qquad (A.49)$$

unconditionally. Fix $\epsilon \in (0, 1)$. Then we may choose some $M > 0$ such that

$$P(\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p > M) \le \alpha\epsilon . \qquad (A.50)$$

By the definition of $\hat{c}^*_{1-\alpha}$, if $\hat{c}^*_{1-\alpha} > M$, then we must have

$$P_W(\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p > M) > \alpha . \qquad (A.51)$$

We may then conclude from the implication of (A.51) that

$$\begin{aligned}
P(\hat{c}^*_{1-\alpha} > M) &\le P_X(P_W(\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p > M) > \alpha) \\
&\le \alpha^{-1} E_X[P_W(\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p > M)] \\
&\le \alpha^{-1} P(\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p > M) \le \epsilon , \quad (A.52)
\end{aligned}$$

where the second inequality follows by the Markov's inequality, the third inequality by Lemma 1.2.6 in van der Vaart and Wellner (1996), and the last step by the result (A.50). This shows that $\hat{c}^*_{1-\alpha} = O_p(1)$ and hence $\hat{c}_\alpha = \max\{\kappa_n, \hat{c}^*_{1-\alpha}\} = O_p(1)$ in view of $\kappa_n = o(1)$ by Assumption 3.1(ii).

Next, by the triangle inequality and Lemma 2.2 in Durot and Tocquet (2003),

$$\begin{aligned}
\sqrt{n}\phi(\mathbb{F}_n) = \sqrt{n}\|\mathcal{M}\mathbb{F}_n - \mathcal{M}F - \{\mathbb{F}_n - F\} + \mathcal{M}F - F\|_p \\
\ge \sqrt{n}\|\mathcal{M}F - F\|_p - \{\sqrt{n}\|\mathcal{M}\mathbb{F}_n - \mathcal{M}F\|_p + \sqrt{n}\|\mathbb{F}_n - F\|_p\} \\
\ge \sqrt{n}\|\mathcal{M}F - F\|_p - C\|\sqrt{n}\{\mathbb{F}_n - F\}\|_\infty , \qquad (A.53)
\end{aligned}$$

where $C > 0$ is a constant depending on the weighting function g. Note that if F is noncacave, then $\mathcal{M}F$ and F must differ from each other on a set of positive Lebesgue measure, implying that $\|\mathcal{M}F - F\|_p > 0$. Together with $\|\sqrt{n}\{\mathbb{F}_n - F\}\|_\infty = O_p(1)$, we may then conclude from result (A.53) that $\sqrt{n}\phi(\mathbb{F}_n) \to \infty$ in probability. Combining with $\hat{c}_\alpha = O_p(1)$, it follows that

$$P(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_\alpha) \to 1 \ . \tag{A.54}$$

This completes the proof for the second claim of the theorem. $\qquad\qquad\square$

Proof of Theorem 3.2. First, for convenience of the reader, we introduce the concept of contiguity. For each $n \in \mathbf{N}$, let P_n and Q_n be two generic probability measures on some measurable space $(\Omega_n, \mathcal{A}_n)$. Then the sequence $\{Q_n\}$ is said to be contiguous with respect to $\{P_n\}$ if, for any statistic $T_n : \Omega_n \to \mathbf{R}$, one has $T_n \xrightarrow{p} 0$ under Q_n whenever $T_n \xrightarrow{p} 0$ under P_n. Heuristically, this means that any statistic T_n that is asymptotically negligible under P_n remains so under Q_n. We say that $\{Q_n\}$ and $\{P_n\}$ are mutually contiguous if $\{Q_n\}$ is contiguous with respect to $\{P_n\}$ and vice versa. We refer the reader to Chapter 6 in van der Vaart (1998)—in particular, Lemma 6.4 there—for more details.

Since $\{P_t\}$ is a differentiable path, it follows by Theorem 12.2.3 and Corollary 12.3.1 in Lehmann and Romano (2005) that P_n^n and P^n are mutually contiguous. By Theorem 2.1 in Fang and Santos (2019) and $\phi(F(P)) = 0$, we then have:

$$\sqrt{n}\phi(\mathbb{F}_n) = \sqrt{n}\{\phi(\mathbb{F}_n) - \phi(F(P))\}$$
$$= \phi'_F(\sqrt{n}\{\mathbb{F}_n - F(P)\}) + o_p(1) \text{ under } P^n \ , \tag{A.55}$$

where $\phi'_F(h) = \|\mathcal{M}'_F h - h\|_p$. In turn, we obtain by (A.55) and mutual continuity of P_n^n and P^n that, under P_n^n,

$$\sqrt{n}\phi(\mathbb{F}_n) = \phi'_F(\sqrt{n}\{\mathbb{F}_n - F(P)\}) + o_p(1) \ . \tag{A.56}$$

As is well known in the literature (see, for example, Bickel et al. (1998, p.192)), $\mathbb{F}_n$ is a regular estimator of F so that, under P_n^n,

$$\sqrt{n}\{\mathbb{F}_n - F(P_{\eta/\sqrt{n}})\} \xrightarrow{L} \mathbb{G} \text{ in } \ell^\infty(\mathbf{R}_+) \ . \tag{A.57}$$

By result (2.4), we also have, as a deterministic result,

$$\sqrt{n}\{F(P_{\eta/\sqrt{n}}) - F(P)\} \to \eta\dot{F}(h) \text{ in } \ell^\infty(\mathbf{R}_+) \ . \tag{A.58}$$

By Slutsky's theorem, we obtain from results (A.57) and (A.58) that

$$\sqrt{n}\{\mathbb{F}_n - F(P)\} \xrightarrow{L} \mathbb{G} + \eta\dot{F}(h) \text{ in } \ell^\infty(\mathbf{R}_+) \ , \tag{A.59}$$

under P_n^n. Combining (A.56) and (A.59), together with the continuous mapping theorem and Slutsky's theorem, we may thus conclude that, under P_n^n,

$$\sqrt{n}\phi(\mathbb{F}_n) \xrightarrow{L} \phi'_F(\mathbb{G} + \eta\dot{F}(h)) \equiv \|\mathcal{M}'_F(\mathbb{G} + \eta\dot{F}(h)) - \{\mathbb{G} + \eta\dot{F}(h)\}\|_p \ . \tag{A.60}$$

For the first part of the theorem, we consider the two cases separately.

<u>Case I:</u> F is strictly concave. By Assumption 3.1(ii), $n^{-2/3}(\log n)^{2/3}/(n^{-1/2}\kappa_n)$ $\to 0$ as $n \to \infty$. By Assumption 3.1(ii) and Theorem 2.2 we thus have

$$\liminf_{n\to\infty} \pi_n(P_{1/\sqrt{n}}) = \liminf_{n\to\infty} P_n^n(\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_p > \max\{\kappa_n, \hat{c}_{1-\alpha}\})$$

$$\leq \liminf_{n\to\infty} P_n^n(\frac{\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_p}{\kappa_n} > 1) = 0 \ . \quad \text{(A.61)}$$

<u>Case II:</u> F is non-strictly concave. Mutual contiguity of P_n^n and P^n, together with $\hat{c}_{1-\alpha}^* \xrightarrow{p} c_{1-\alpha}^*$ under P^n from the proof of Theorem 3.1, implies that whenever F is non-strictly concave,

$$\hat{c}_{1-\alpha}^* \xrightarrow{p} c_{1-\alpha}^* \text{ under } P_n^n \ . \quad \text{(A.62)}$$

Since $c_{1-\alpha}^*$ is nonnegative as a quantile of $\|\mathcal{M}_F'(\mathbb{G}) - \mathbb{G}\|_p$ and $\kappa_n = o(1)$ by Assumption 3.1(ii), we obtain from (A.62) and the continuous mapping theorem that $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}^*$ under P_n^n. This, together with result (A.60), Slutsky's theorem and the portmanteau theorem, allows us to conclude that

$$\liminf_{n\to\infty} \pi_n(P_{1/\sqrt{n}}) = \liminf_{n\to\infty} P_n^n(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{1-\alpha})$$

$$= P(\|\mathcal{M}_F'(\mathbb{G} + \eta\dot{F}(h)) - \{\mathbb{G} + \eta\dot{F}(h)\}\|_p > c_{1-\alpha}^*) \ . \quad \text{(A.63)}$$

If in addition the cdf of $\|\mathcal{M}_F'(\mathbb{G} + \eta\dot{F}(h)) - \{\mathbb{G} + \eta\dot{F}(h)\}\|_p$ is continuous at $c_{1-\alpha}^*$, then (A.63) holds with equality, again by the portmanteau theorem. This completes the proof for the first part of the theorem.

As for the second part, we again consider the following two cases separately.

<u>Case I:</u> F is strictly concave. By Assumption 3.1(ii) and Theorem 2.2 we have

$$\limsup_{n\to\infty} \pi_n(P_{1/\sqrt{n}}) \leq \limsup_{n\to\infty} P_n^n(\sqrt{n}\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_p > \kappa_n) = 0 \ . \quad \text{(A.64)}$$

This is in fact also implied by result (A.61) which holds for any $\eta \in \mathbf{R}$.

<u>Case II:</u> F is non-strictly concave. We follow the arguments of Theorem 3.3 in Fang and Santos (2019). To begin with, note that result (A.60), $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}^*$ under P_n^n (as argued previously), Slutsky's theorem and the portmanteau theorem together imply that

$$\limsup_{n\to\infty} \pi_n(P_{1/\sqrt{n}}) \leq \limsup_{n\to\infty} P_n^n(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{1-\alpha}^*)$$

$$\leq P(\|\mathcal{M}_F'(\mathbb{G} + \eta\dot{F}(h)) - \{\mathbb{G} + \eta\dot{F}(h)\}\|_p \geq c_{1-\alpha}^*) \ . \quad \text{(A.65)}$$

Next, we show that the map $g \mapsto \phi_F'(g) = \|\mathcal{M}_F'(g) - g\|_p$ (defined on $\ell^\infty(\mathbf{R}_+)$) is subadditive. Note that (i) $\mathcal{M}_F'(g) - g \geq 0$ for all $g \in \ell^\infty(\mathbf{R}_+)$ because $\mathcal{M}_F'(g)$

majorizes g, and (ii) $\ell^\infty(\mathbf{R}_+)$ can be identified as a subspace of $L^p(\mathbf{R}_+)$ equipped with norm $\|\cdot\|_p$ if $p \in [1,\infty)$—of course, if $p = \infty$, then $\ell^\infty(\mathbf{R}_+)$ is a subspace of itself. Thus, we may view ψ'_F as a real-valued map defined on some (normed) subspace of $L^p(\mathbf{R}_+)$ if $p \in [1,\infty)$ or of $\ell^\infty(\mathbf{R}_+)$ if $p = \infty$. Since ϕ'_F is defined in term of the $\|\cdot\|_p$ norm, such an embedding (of the domain $\ell^\infty(\mathbf{R}_+)$ of ϕ'_F into $L^p(\mathbf{R}_+)$) allows us to rewrite ϕ'_F according to Lemma A.1.

Specifically, let $\ell^\infty(\mathbf{R}_+)^{p*}$ be the topological dual space of $\ell^\infty(\mathbf{R}_+)$ viewed as a subspace of $L^p(\mathbf{R}_+)$ if $p \in [1,\infty)$, and be the topological dual space of $\ell^\infty(\mathbf{R}_+)$ if $p = \infty$.[3] Define $\mathbb{S}^p_+ = \{\varphi \in \ell^\infty(\mathbf{R}_+)^{p*} : \|\varphi\|_{op} = 1, \varphi \geq 0\}$ where $\|\cdot\|_{op}$ is the operator norm, and $\varphi \geq 0$ means that $\varphi(g) \geq 0$ whenever $g(x) \geq 0$ for all $x \in \mathbf{R}_+$. By Lemma A.1 we now have: for any $g_1, g_2 \in \ell^\infty(\mathbf{R}_+)$,

$$\phi'_F(g_1 + g_2) = \sup_{\varphi \in \mathbb{S}^p_+} \varphi(\mathcal{M}'_F(g_1 + g_2) - (g_1 + g_2))$$

$$= \sup_{\varphi \in \mathbb{S}^p_+} \left\{ \varphi(\mathcal{M}'_F(g_1 + g_2)) - [\varphi(g_1) + \varphi(g_2)] \right\} , \quad \text{(A.66)}$$

where the second equality follows by linearity of each $\varphi \in \mathbb{S}^p_+$. Since $\mathcal{M}'_F$ is convex by Proposition 2.2 in Beare and Fang (2017) and positively homogeneous of degree one as a Hadamard directional derivative (Shapiro, 1990), it follows that $\mathcal{M}'_F$ must be subadditive, i.e.,

$$\mathcal{M}'_F(g_1 + g_2) \leq \mathcal{M}'_F(g_1) + \mathcal{M}'_F(g_2) . \quad \text{(A.67)}$$

Since each $\varphi \in \mathbb{S}^p_+$ is linear and satisfies $\varphi(g) \geq 0$ whenever $g(x) \geq 0$ for all $x \in \mathbf{R}_+$, we obtain from (A.67) that, for any $\varphi \in \mathbb{S}^p_+$,

$$\varphi(\mathcal{M}'_F(g_1) + \mathcal{M}'_F(g_2)) \leq \varphi(\mathcal{M}'_F(g_1)) + \varphi(\mathcal{M}'_F(g_2)) . \quad \text{(A.68)}$$

Combining results (A.66) and (A.68) then leads to

$$\phi'_F(g_1 + g_2) \leq \sup_{\varphi \in \mathbb{S}^p_+} \left\{ \varphi(\mathcal{M}'_F(g_1)) + \varphi(\mathcal{M}'_F(g_2)) - [\varphi(g_1) + \varphi(g_2)] \right\}$$

$$\leq \sup_{\varphi \in \mathbb{S}^p_+} \left\{ \varphi(\mathcal{M}'_F(g_1)) - \varphi(g_1) \right\} + \sup_{\varphi \in \mathbb{S}^p_+} \left\{ \varphi(\mathcal{M}'_F(g_2)) - \varphi(g_2) \right\}$$

$$= \sup_{\varphi \in \mathbb{S}^p_+} \left\{ \varphi(\mathcal{M}'_F(g_1) - g_1) \right\} + \sup_{\varphi \in \mathbb{S}^p_+} \left\{ \varphi(\mathcal{M}'_F(g_2) - g_2) \right\} , \quad \text{(A.69)}$$

where the equality again follows by linearity of each $\varphi \in \mathbb{S}^p_+$. By Lemma A.1, we in turn have from (A.69) that, for any $g_1, g_2 \in \ell^\infty(\mathbf{R}_+)$,

$$\phi'_F(g_1 + g_2) \leq \|\mathcal{M}'_F(g_1) - g_1\|_p + \|\mathcal{M}'_F(g_1) - g_1\|_p$$

$$= \phi'_F(g_1) + \phi'_F(g_2) . \quad \text{(A.70)}$$

---

[3]Let $\mathbb{D}$ be a normed space with norm $\|\cdot\|_{\mathbb{D}}$ such as $L^p(\mathbf{R}_+)$ and $\ell^\infty(\mathbf{R}_+)$. The topological dual space $\mathbb{D}^*$ of $\mathbb{D}$ is the space of all continuous and linear functions $\varphi : \mathbb{D} \to \mathbf{R}$, equipped with the operator norm $\|\varphi\|_{op} \equiv \sup_{\|g\|_{\mathbb{D}} \leq 1} |\varphi(g)|$.

In light of result (A.66), we immediately obtain that

$$
P(\|\mathcal{M}'_F(\mathbb{G} + \eta\dot{F}(h)) - \{\mathbb{G} + \eta\dot{F}(h)\}\|_p \geq c^*_{1-\alpha})
$$
$$
\leq P(\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p + \|\mathcal{M}'_F(\eta\dot{F}(h)) - \eta\dot{F}(h)\|_p \geq c^*_{1-\alpha}) . \quad \text{(A.71)}
$$

Moreover, since $\{P_{\eta/\sqrt{n}}\}$ is a differentiable path under the null when $\eta \leq 0$, it follows that $\phi(F(P_{\eta/\sqrt{n}})) = \phi(F(P)) = 0$ for all $n \in \mathbf{N}$ and $\eta \leq 0$. Hence,

$$
0 = \lim_{n\to\infty} \frac{\phi(F(P_{\eta/\sqrt{n}})) - \phi(F(P))}{n^{-1/2}}
$$
$$
= \phi'_F(\eta\dot{F}(h)) \equiv \|\mathcal{M}'_F(\eta\dot{F}(h)) - \eta\dot{F}(h)\|_p . \quad \text{(A.72)}
$$

Results (A.65), (A.71) and (A.72) then yield that

$$
\limsup_{n\to\infty} \pi_n(P_{1/\sqrt{n}}) \leq P(\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p \geq c^*_{1-\alpha}) = \alpha ,
$$

where we also exploited that $c^*_{1-\alpha}$ is a continuity point of the cdf of $\|\mathcal{M}'_F(\mathbb{G}) - \mathbb{G}\|_p$ by Assumption 3.2. This completes the proof of the second part.

Finally, for the third part of the theorem, we note by Lemma 2.2 in Durot and Tocquet (2003) that

$$
|\sqrt{n}\phi(\mathbb{F}_n) - \sqrt{n}\phi(F_n)| \lesssim \|\sqrt{n}\{\mathbb{F}_n - F_n\}\|_p \lesssim \sqrt{n}\|\mathbb{F}_n - F_n\|_\infty , \quad \text{(A.73)}
$$

where the second inequality follows from the definition of $\|\cdot\|_p$. Fix $M > 0$. By Dvoretzky, Kiefer and Wolfowitz's inequality (see, for example, Theorem 11.2.18 in Lehmann and Romano (2005)), we have that

$$
P_n^n(\sqrt{n}\|\mathbb{F}_n - F_n\|_\infty > M) = P_n(\|\mathbb{F}_n - F_n\|_\infty > \frac{M}{\sqrt{n}}) \lesssim \exp\{-2M^2\} , \quad \text{(A.74)}
$$

which can be made arbitrarily small by choosing a sufficiently large M. Combining results (A.73) and (A.74) we thus obtain that

$$
|\sqrt{n}\phi(\mathbb{F}_n) - \sqrt{n}\phi(F_n)| = O_p(1) \text{ under } P_n^n . \quad \text{(A.75)}
$$

Moreover, by Proposition 2.1 in Beare and Fang (2017) and $\phi(F) = 0$ (by the definition of **H**), we obtain that, as $n \to \infty$,

$$
\sqrt{n}\phi(F_n) = \sqrt{n}\{\phi(F_n) - \phi(F)\} \to \|\mathcal{M}'_F(\eta\dot{F}(h)) - \eta\dot{F}(h)\|_p . \quad \text{(A.76)}
$$

By the positive homogeneity of degree one of $\mathcal{M}'_F$ (as a Hadamard directional derivative), we thus have from (A.76) that: for $\Delta \equiv \|\mathcal{M}'_F(\dot{F}(h)) - \dot{F}(h)\|_p$,

$$
\sqrt{n}\phi(F_n) = \eta\Delta + o(1) , \quad \text{(A.77)}
$$

for all $\eta > 0$. It follows from (A.75) and (A.77) that, under P_n^n,

$$
\sqrt{n}\phi(\mathbb{F}_n) = \sqrt{n}\phi(\mathbb{F}_n) - \sqrt{n}\phi(F_n) + \sqrt{n}\phi(F_n) = \eta\Delta + O_p(1) . \quad \text{(A.78)}
$$

Having derived the order of the test statistic as in (A.78), we next evaluate the order of the critical value. For this, note that, by result (A.48),

$$\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}) - \sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}\|_p \lesssim 2\sqrt{n}\|\mathbb{F}^*_n - \mathbb{F}_n\|_\infty . \tag{A.79}$$

By Dvoretzky, Kiefer and Wolfowitz's inequality (see, for example, Theorem 11.2.18 in Lehmann and Romano (2005)), we note that, for any $M > 0$,

$$P_W(\sqrt{n}\|\mathbb{F}^*_n - \mathbb{F}_n\|_\infty > M) \lesssim \exp\{-2M^2\} , \tag{A.80}$$

which can be made arbitrarily small by choosing a sufficiently large M. Result (A.80), together with arguments similar to those leading to (A.52), yields

$$\hat{c}_{1-\alpha^*} = O_p(1) \text{ under } P^n_n . \tag{A.81}$$

Combining results (A.78) and (A.81), we may thus conclude

$$P^n_n(\sqrt{n}\phi(\mathbb{F}_n) > \hat{c}_{n,1-\alpha}) = P^n_n(\eta\Delta + O_p(1) > \max\{\kappa_n, \hat{c}^*_{1-\alpha}\}) = 1 , \tag{A.82}$$

by letting $n \to \infty$ followed by $\eta \uparrow \infty$, where we exploited $\kappa_n = o(1)$ by Assumption 3.1(ii) and $\Delta > 0$. This completes the proof of the theorem. □

Our final lemma entails some knowledge on Riesz space. For convenience of the reader, we introduce some relevant concepts in this regard. Just like normed spaces generalize to abstract spaces the operations of vector addition and scalar multiplication in Euclidean spaces, Riesz spaces generalize the binary relations $\geq$ and $\leq$. Let $\mathbb{E}$ be a set. We say that $\mathbb{E}$ is partially ordered if there exists a binary relation $\geq$ that is transitive (i.e., $x \geq z$ whenever $x \geq y$ and $y \geq z$), reflexive (i.e., $x \geq x$) and antisymetric (i.e., $x \geq y$ and $y \geq x$ implies $x = y$). The notation $y \leq x$ is equivalent to $x \geq y$. If $x \geq y$ but $x \neq y$, we also write $x > y$ or equivalently $y < x$. Let $\mathbb{E}$ be partially ordered by $\geq$. An element z is the supremum of a pair $x, y \in \mathbb{E}$, denoted $x \vee y$, if $x \geq z$, $y \geq z$, and $z \leq u$ whenever $x \leq u$ and $y \leq u$. The infimum of $x, y \in \mathbb{E}$, denoted $x \wedge y$, is defined similarly. Not every pair x, y in $\mathbb{E}$ admits a supremum or infimum. But, if this is the case, then we call $\mathbb{E}$ a lattice.

If $\mathbb{E}$ is a vector space, then one may hope that the partial order $\geq$ is "compatible" with the algebraic structure of $\mathbb{E}$. Concretely, whenever $x \geq y$, one should have (i) $x + z \geq y + z$ for all $z \in \mathbb{E}$, and (ii) $\alpha x \geq \alpha y$ for all $\alpha \in \mathbf{R}_+$. In this case, we call $\mathbb{E}$ an (partially) ordered vector space. A partially ordered vector space that is also a lattice is called a Riesz space. For example, for $L^p(\mathbf{R}_+)$ with $p \in [1, \infty)$, we may define the partial order $\geq$ by: $f \geq g$ whenever $f(x) \geq g(x)$ for almost all $x \in \mathbf{R}_+$. For $\ell^\infty(\mathbf{R}_+)$, we define $\geq$ analogously: $f \geq g$ whenever $f(x) \geq g(x)$ for all $x \in \mathbf{R}_+$. Then both $L^p(\mathbf{R}^+)$ and $\ell^\infty(\mathbf{R}_+)$ are Riesz spaces. Given the notion of supremum, we may define the absolute value of $x \in \mathbb{E}$ by $|x| \equiv x^+ + x^-$ where $x^+ \equiv x \vee 0$ and $x^- \equiv (-x) \vee 0$. If $\mathbb{E}$ is a Riesz space and is equipped with a norm $\|\cdot\|$ satisfying the property that $\|x\| \leq \|y\|$ whenever $|x| \leq |y|$, then $\mathbb{E}$ is called a normed Riesz space, and the norm $\|\cdot\|$ is called a lattice norm. We refer the reader to Aliprantis and Border (2006) for more discussions. With these concepts in hand, we now have:

Lemma A.1. *Let $\mathbb{E}$ be a normed Riesz space with partial order $\geq$ and lattice norm $\|\cdot\|_{\mathbb{E}}$. Then for any $x \geq 0$, we have*

$$\|x\|_{\mathbb{E}} = \sup_{\varphi \in \mathbb{E}^* : \|\varphi\|_{op}=1, \varphi \geq 0} \varphi(x) \ ,$$

where $\mathbb{E}^*$ is the topological dual space of $\mathbb{E}$, $\|\cdot\|_{op}$ is the operator norm, and $\varphi \geq 0$ means that $\varphi(y) \geq 0$ whenever $y \geq 0$.

Proof. The conclusion trivially holds if $x = 0$. Suppose that $x > 0$. Then Theorem 39.3 in Zaanen (1997) implies that there exists some $\varphi^* \in \mathbb{E}^*$ with $\|\varphi^*\|_{op} = 1$ and $\varphi^* \geq 0$ such that $\varphi^*(x) = \|x\|_{\mathbb{E}}$. This in turn implies that

$$\|x\|_{\mathbb{E}} = \varphi^*(x) \leq \sup_{\varphi \in \mathbb{E}^* : \|\varphi\|_{op}=1, \varphi \geq 0} \varphi(x) \ .$$

On the other hand, we have

$$\sup_{\varphi \in \mathbb{E}^* : \|\varphi\|_{op}=1, \varphi \geq 0} \varphi(x) \leq \sup_{\varphi \in \mathbb{E}^* : \|\varphi\|_{op}=1, \varphi \geq 0} \|\varphi\|_{op}\|x\|_{\mathbb{E}} = \|x\|_{\mathbb{E}} \ .$$

This completes the proof of the lemma. $\qquad\square$

Appendix B: Some extensions

In this section, we discuss extensions of our test results to a general setup that includes inference on density, regression and hazard functions as special cases. To this end, We shall assume that the data $\{X_i\}_{i=1}^n$ are i.i.d. with common distribution $P_{\eta/\sqrt{n}} \in \mathbf{P}$, which we also denote by P_n with some abuse of notation. Here, $\mathbf{P}$ denotes the model as before. This allows us to consider the standard i.i.d. setup (i.e., $\eta = 0$) as well as the setup for local analysis, in a unified way. In turn, wherever appropriate, we identify the parameter of interest θ as a map $\theta : \mathbf{P} \to \ell^\infty([a, b])$ such that, for any $P \in \mathbf{P}$ and $t \in [a, b]$,

$$\theta(P)(t) = \int_a^t g(u)\, \mathrm{d}u \ , \tag{B.1}$$

where $a, b \in \mathbf{R}$ are known with $a < b$ throughout. The dependence of $\theta(P)$ on P sometimes is also suppressed for simplicity. While it may be possible to consider settings with unbounded support such as $[0, \infty)$, we confine our attention to the bounded case for simplicity—it is not essential to what we intend to convey anyway. Our objective is to test the null that $\theta(P_{\eta/\sqrt{n}})$ is concave versus otherwise. In the main text, g is the density function and θ is the cdf. Below we formalize the hazard and the regression examples.

Example B.1 (Monotone Hazard Rate)**.** Let $X \in \mathbf{R}_+$ be a random variable with cdf F that admits a pdf f with respect to the Lebesgue measure. Then the hazard rate λ at u is defined by:

$$\lambda(u) \equiv \frac{f(u)}{1 - F(u)} \ . \tag{B.2}$$

Heuristically, $\lambda(u)$ measures the probability of instantaneous failure rate at time u, given the subject has functioned until u (Groeneboom and Jongbloed, 2014). A leading example of shape restrictions is that λ be monotonically increasing on an interval $[0, b]$ for some $b \in (0, \infty)$. This can be inferred from convexity of the cumulative hazard function Λ defined by: for any $t \in [0, b]$,

$$\Lambda(t) \equiv \Lambda_F(t) \equiv \int_0^t \frac{1}{1 - F_-(u)} \, \mathrm{d}F(u) , \tag{B.3}$$

where $u \mapsto F_-(u) \equiv \lim_{t \uparrow u} F(t)$ is the left-continuous version of F. Here, $F(u) = F(u-)$ for all $u \in \mathbf{R}_+$ since F is absolutely continuous, so $\Lambda(t) = \int_0^t \lambda(u) \, \mathrm{d}u$. The definition above, however, allows us to construct estimators of Λ in a straightforward way. In this example, $\theta = -\Lambda$. $\qquad\qquad\qquad\qquad\qquad\square$

Example B.2 (Isotonic/Monotone Regression). Let $\{(Y_i, Z_i)\}_{i=1}^n$ be i.i.d. bivariate random vectors generated according to

$$Y_i = m(Z_i) + \epsilon_i , \tag{B.4}$$

where each Z_i is supported on $[a, b]$, the regression function $m : [a, b] \to \mathbf{R}$ is unknown, and each ϵ_i is a centered random variable independent of Z_i. Let $Q : [0, 1] \to [a, b]$ be the population quantile function of Z_1. Then monotonicity of m can be inferred from convexity of the population cumulative regression function (cqf) $\theta \in \ell^\infty([0, 1])$ defined by: for all $t \in [0, 1]$,

$$\Lambda(t) = \int_0^t m(Q(u)) \, \mathrm{d}u . \tag{B.5}$$

One may also view Λ as a population analog of the cumulative sum diagram—see, for example, Beare and Fang (2017, p.3865). Here, $\theta = -\Lambda$. $\qquad\quad\square$

Since a function $\theta : [a, b] \to \mathbf{R}$ is convex if and only if $-\theta$ is concave, both Examples B.2 and B.1 fall within the scope of our framework. For a coherent treatment, we shall thus maintain the null hypothesis that θ is concave versus otherwise, and revisit Examples B.2 and B.1 after the general theory is presented. Towards this end, let $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \ell^\infty([a, b])$ be an unconstrained estimator of θ. Note that X_i is a generic notation for the i-th observation; e.g., $X_i = (Y_i, Z_i)'$ in Example B.2. In turn, we then employ the test statistic $\sqrt{n}\phi(\hat{\theta}_n)$ where $\phi : \ell^\infty([a, b]) \to \mathbf{R}$ is defined by

$$\phi(\theta) = \|\mathcal{M}\theta - \theta\|_p = \begin{cases} [\int_a^b (\mathcal{M}\theta(x) - \theta(x))^p g(x) \, \mathrm{d}x]^{1/p} & \text{if } p \in [1, \infty) \\ \sup_{a \le x \le b} |\mathcal{M}\theta(x) - \theta(x)| & \text{if } p = \infty \end{cases} ,$$

for some positive weighting function $g \in L^1([a, b])$. Throughout, we study the test functional ϕ with an arbitrarily fixed $p \in [1, \infty]$.

We now proceed by imposing the following assumptions, where $C([a,b])$ is the space of continuous functions on $[a,b]$ equipped with the uniform norm.

Assumption B.1. (i) $\{X_i\}_{i=1}^n$ are i.i.d. with a common distribution $P_{\eta/\sqrt{n}}$ for some $\eta \in \mathbf{R}$; (ii) $\{P_t : |t| < \epsilon\} \subset \mathbf{P}$ is a differentiable path with score h in the sense of Definition (2.1); (iii) $\|\sqrt{n}\{\theta(P_{\eta/\sqrt{n}}) - \theta(P_0)\} - \eta\dot{\theta}(h)\|_\infty \to 0$ for some $\eta\dot{\theta}(h) \in C([a,b])$; (iv) $\theta_0 \equiv \theta(P_0) \in \ell^\infty([a,b])$ is concave.

Assumption B.2. (i) $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \ell^\infty([a,b])$ satisfies $\sqrt{n}\{\hat{\theta}_n - \theta_n\} \xrightarrow{L} \mathbb{G}$ in $\ell^\infty([a,b])$ under $P_n^n \equiv \prod_{i=1}^n P_{\eta/\sqrt{n}}$, where $\theta_n \equiv \theta(P_{\eta/\sqrt{n}})$ and $\mathbb{G} \in C([a,b])$; (ii) $\|\mathcal{M}\hat{\theta}_n - \hat{\theta}_n\|_\infty = O_p(s_n)$ under P_n^n if θ_0 is strictly concave for some sequence $\{s_n\}$ of strictly positive scalars satisfying $\sqrt{n}s_n = o(1)$.

Assumption B.1(i)(ii) formalizes the data generating process. Assumption B.1(iii) is a generalization of the property (2.4), and is fulfilled whenever $\theta : \mathbf{P} \to \ell^\infty([a,b])$ is a regular parameter in the sense of Definition 5.1 in Bickel et al. (1998). If θ_n and θ_0 are in $C([a,b])$ for all n, then so is $\eta\dot{\theta}(h)$ as a uniform limit of continuous functions. Assumption B.1(iv) implies that $\{P_{\eta/\sqrt{n}}\}$ is a sequence of local perturbations such that θ_n tends to a concave function θ_0. We stress that the parameter of interest in truth is θ_n while θ_0 is the limit of $\{\theta_n\}$; if $\eta = 0$ (e.g., the i.i.d. setup with a fixed distribution P_0), then $\theta_n = \theta_0$ (and $\eta\dot{\theta}(h) = 0$). Assumption B.2(i) requires an estimator $\hat{\theta}_n$ of θ_n that admits a weak limit $\mathbb{G}$ at rate $\sqrt{n}$. In turn, Assumption B.2(ii) is a high level condition describing the consequence of degeneracy when θ_0 is strictly concave. For the density problem in the main text, we have $s_n = (n^{-1}\log n)^{2/3}$ as verified by Theorem 2.2. Note that the order in probability (i.e., O_p) instead of almost surely (i.e., $O_{a.s.}$) suffices for our inferential purpose.

By a simple generalization of Lemma 3.2 in Beare and Moon (2015), $\mathcal{M} : \ell^\infty([a,b]) \to \ell^\infty([a,b])$ is Hadamard directionally differentiable at the concave θ_0 tangentially to $C([a,b])$. Like $\mathcal{M}'_F$ in the main text, the derivative $\mathcal{M}'_{\theta_0}(h)$ majorizes h by concave functions on regions over which θ_0 is affine but acts like an identity map elsewhere. We omit the explicit expression of the derivative $\mathcal{M}'_{\theta_0}$ as it is not important to the implementation of the bootstrap, but refer the reader to Beare and Moon (2015). By Assumptions B.1 and B.2(i), we in turn have by analogous arguments as in the proof of Theorem 3.2 that

$$\sqrt{n}\phi(\hat{\theta}_n) \xrightarrow{L} \|\mathcal{M}'_{\theta_0}(\mathbb{G} + \eta\dot{\theta}(h)) - (\mathbb{G} + \eta\dot{\theta}(h))\|_p \text{ under } P_n^n . \tag{B.6}$$

Therefore, letting $\eta = 0$ yields the pointwise weak limit

$$\|\mathcal{M}'_{\theta_0}(\mathbb{G}) - (\mathbb{G})\|_p . \tag{B.7}$$

If θ_0 is strictly concave, then $\mathcal{M}'_{\theta_0}$ is the identity map so that (B.7) is degenerate at 0, consistent with Assumption B.2(ii). Following the main text, we differentiate strict concavity from non-strict concavity by introducing a tuning parameter $\kappa_n > 0$ such that $\kappa_n \to 0$ and $\sqrt{n}s_n/\kappa_n \to 0$ as $n \to \infty$.

Towards construction of the critical values, we estimate the law of (B.7) by the rescaled bootstrap. First, we estimate $\mathcal{M}'_{\theta_0}$ by: for any $h \in \ell^\infty([a,b])$,

$$\hat{\mathcal{M}}'_n(h) = \frac{\mathcal{M}(\hat{\theta}_n + t_n h) - \mathcal{M}(\hat{\theta}_n)}{t_n} \ , \tag{B.8}$$

where $t_n \downarrow 0$ such that $t_n \sqrt{n} \to \infty$. Next, we need to bootstrap the law of $\mathbb{G}$. In order to accommodate flexible bootstrap schemes, we consider a general bootstrap quantity $\hat{\mathbb{G}}_n : \{X_i, W_{ni}\}_{i=1}^n \to \ell^\infty([a,b])$ where $\{W_{ni}\}_{i=1}^n$ are bootstrap weights independent of the data $\{X_i\}_{i=1}^n$. For example, for the standard empirical bootstrap, $W_n \equiv (W_{n1}, \ldots, W_{nn})$ is a multinomial random vector over n categories with probabilities $(1/n, \ldots, 1/n)$. More generally, one may consider multiplier or exchangeable bootstrap, corresponding respectively to i.i.d. or exchangeable weights (Praestgaard and Wellner, 1993; Kosorok, 2008). To formalize the notion of bootstrap consistency, we employ the bounded Lipschitz metric (van der Vaart and Wellner, 1996). For a generic normed space $\mathbb{D}$ equipped with norm $\|\cdot\|_\mathbb{D}$ (e.g., $\mathbb{D} = \ell^\infty([a,b])$), let

$$\mathrm{BL}_1(\mathbb{D}) \equiv \{f : \mathbb{D} \to \mathbf{R} : \sup_{x \in \mathbb{D}} |f(x)| \leq 1 \text{ and } \sup_{x,y \in \mathbb{D}: x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|_\mathbb{D}} \leq 1\} \ .$$

If $\hat{\mathbb{G}}_n : \{X_i, W_{ni}\}_{i=1}^n \to \mathbb{D}$ is a bootstrap for the law of a random element $\mathbb{G}$ in $\mathbb{D}$, then $\hat{\mathbb{G}}_n$ is said to be a consistent bootstrap for $\mathbb{G}$ if

$$\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(\hat{\mathbb{G}}_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G})]| = o_p(1) \ . \tag{B.9}$$

Finally, for $\alpha \in (0,1)$, we set our critical value $\hat{c}_{1-\alpha} = \max\{\kappa_n, \hat{c}^*_{1-\alpha}\}$ where

$$\hat{c}^*_{1-\alpha} \equiv \inf\{c \in \mathbf{R} : P(\|\hat{\mathcal{M}}'_n(\hat{\mathbb{G}}_n) - (\hat{\mathbb{G}}_n)\|_p \leq c|\{X_i\}_{i=1}^n) \geq 1 - \alpha\} \ . \tag{B.10}$$

We then reject the null that θ_n is concave if $\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}$.

To ensure validity of $\hat{c}_{1-\alpha}$ as our critical value, we impose:

Assumption B.3. (i) $\hat{\mathbb{G}}_n : \{X_i, W_{ni}\}_{i=1}^n \to \ell^\infty([a,b])$ where $\{W_{ni}\}_{i=1}^n$ are bootstrap weights independent of $\{X_i\}_{i=1}^n$; (ii) $\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(\hat{\mathbb{G}}_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G})]| = o_p(1)$; (iii) $\hat{\mathbb{G}}_n$ is asymptotically measurable in $\{X_i, W_{ni}\}_{i=1}^n$; (iv) $f(\hat{\mathbb{G}}_n)$ is a measurable function of $\{W_{ni}\}_{i=1}^n$ outer almost surely in $\{X_i\}_{i=1}^n$ for any continuous and uniformly bounded $f : \ell^\infty([a,b]) \to \mathbf{R}$.

Assumption B.4. (i) $\{t_n\}$ is a sequence of positive scalars such that $t_n \to 0$ and $\sqrt{n}t_n \to \infty$ as $n \to \infty$; (ii) $\{\kappa_n\}$ is a sequence of positive scalars such that $\kappa_n \to 0$ and $\sqrt{n}s_n/\kappa_n \to 0$ as $n \to \infty$.

Assumption B.5. The cdf of $\|\mathcal{M}'_{\theta_0}(\mathbb{G}) - \mathbb{G}\|_p$ is continuous and strictly increasing at its $(1-\alpha)$th quantile $c^*_{1-\alpha}$ when θ_0 is non-strictly concave.

Assumption B.3(i)(ii) formalize the bootstrap consistency of $\hat{\mathbb{G}}_n$ per our discussions above. Assumption B.3(iii)(iv) are mild technical conditions. The precise meaning of Assumption B.3(iii) is that, for any $f \in \mathrm{BL}_1(\mathbb{D})$,

$$E^*[f(\hat{\mathbb{G}}_n)] - E_*[f(\hat{\mathbb{G}}_n)] \to 0 \, ,$$

where E^* and E_* are respectively outer and inner expectations with respect to $\{X_i, W_{ni}\}_{i=1}^n$ (jointly)—see Chapter 1.2 in van der Vaart and Wellner (1996) for more discussions. Assumption B.3(iii) can be verified by appealing to existing results directly; see Theorems 2.6 and 10.4 in Kosorok (2008), or Theorem 3.6.13 in van der Vaart and Wellner (1996) in conjunction with Lemma S.3.9 in Fang and Santos (2019). Assumption B.3(iv) is met whenever $f(\hat{\mathbb{G}}_n)$ is continuous in $\{W_{ni}\}_{i=1}^n$, as in common bootstrap schemes. In the main text, Assumption B.3 is automatically satisfied for $\hat{\mathbb{G}}_n$ that is constructed by the classical empirical bootstrap—see Theorem 3.6.1 in van der Vaart and Wellner (1996). Assumption B.4 imposes suitable rate conditions on the tuning parameter t_n (for the rescaled bootstrap) and on κ_n (for the KW-selection). Finally, Assumption B.5 is an analog of Assumption 3.2, which ensures that bootstrap consistency delivers consistency of the critical value $\hat{c}_{1-\alpha}^*$.

Given the above assumptions, we may obtain the following analog of Theorem 3.2 which subsumes the pointwise results in Theorem 3.1. As before, let

$$\pi_n(P_{\eta/\sqrt{n}}) \equiv P_n^n(\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) \, .$$

Theorem B.1. *Let Assumptions B.1, B.2, B.3, B.4 and B.5 hold, and set $\hat{c}_{1-\alpha} = \max\{\kappa_n, \hat{c}_{1-\alpha}^*\}$ with $\hat{c}_{1-\alpha}^*$ defined as in* (B.10). *Then it follows that*

1. *For any $\eta \in \mathbf{R}$, (i) if θ_0 is strictly concave, then $\liminf_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) = 0$, and (ii) if θ_0 is non-strictly concave, then*

$$\liminf_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}})$$
$$\geq P(\|\mathcal{M}'_{\theta_0}(\mathbb{G} + \eta\dot{\theta}(h)) - \{\mathbb{G} + \eta\dot{\theta}(h)\}\|_p > c_{1-\alpha}^*) \, , \quad \text{(B.11)}$$

   where (B.11) *holds with equality if, in addition, the cdf of $\|\mathcal{M}'_{\theta_0}(\mathbb{G} + \eta\dot{\theta}(h)) - \{\mathbb{G} + \eta\dot{\theta}(h)\}\|_p$ is continuous at $c_{1-\alpha}^*$.*
2. *For any $\eta \leq 0$, we have*

$$\limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) \leq \alpha \, . \quad \text{(B.12)}$$

3. *If $\|\mathcal{M}'_{\theta_0}(\eta\dot{\theta}(h)) - \eta\dot{\theta}(h)\|_p > 0$, then*

$$\liminf_{\eta\uparrow\infty} \liminf_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) = 1 \, . \quad \text{(B.13)}$$

The proof of Theorem B.1 is in complete accord with the proof of Theorem 3.2 given our high level assumptions, and is thus omitted. In what follows, we therefore focus on verifying some of the main assumptions and in particular Assumption B.2(ii) for Examples B.1 and B.2. Assumptions B.1(i)(ii)(iv) and B.5 may be thought of as regulating the data generating process, while Assumption B.4 is simply concerned with choices of tuning parameters.

Example B.1 (cont.). Let $\mathbf{P}$ be the class of distributions on $\mathbf{R}_+$ dominated by the Lebesgue measure. By Example 5.3.5 in Bickel et al. (1998), $\Lambda : \mathbf{P} \to \ell^\infty([0,b])$ is a regular parameter which may be estimated efficiently by the empirical cumulative hazard function $\hat{\Lambda}_n$: for any $t \in [0,b]$,

$$\hat{\Lambda}_n(t) \equiv \Lambda_{\mathbb{F}_n}(t) = \int_0^t \frac{1}{1 - \mathbb{F}_{n,-}(u)} \, d\mathbb{F}_n(u) \ , \tag{B.14}$$

where $\mathbb{F}_n$ is the empirical cdf and $\mathbb{F}_{n,-}$ is the left-continuous version of $\mathbb{F}_n$. Since the map $F \mapsto \Lambda_F$ is Hadamard differentiable (under regularity conditions)—see, for example, Lemma 20.14 in van der Vaart (1998), by the Delta method in conjunction with weak convergence of $\mathbb{F}_n$, one can show that Assumption B.2(i) is satisfied with $\mathbb{G} = \mathbb{W} \circ \chi \in C([a,b])$, where $\mathbb{W}$ is the standard Brownian motion and $\chi(t) \equiv F(t)/(1 - F(t))$ for any $t \in [0,b]$. We refer the reader also to Chapter 6 in Shorack and Wellner (1986) for a detailed treatment of weak convergence of $\hat{\Lambda}_n$. If $\mathbb{F}_n^*$ is the bootstrap empirical distribution as described below (3.6), then the aforementioned Hadamard differentiability implies that $\hat{\mathbb{G}}_n = \sqrt{n}\{\Lambda_{\mathbb{F}_n^*} - \Lambda_{\mathbb{F}_n}\}$ satisfies Assumption B.3 by a combination of Theorems 3.6.1 and 3.9.11 in van der Vaart and Wellner (1996) and Lemma S.3.9 in Fang and Santos (2019). The main condition to verify is now Assumption B.2(ii). In this regard, we believe that, as in Theorem 2.2, it is possible to obtain the same asymptotic order under $\prod_{i=1}^n P_{\eta/\sqrt{n}}$ as under $\prod_{i=1}^n P_0$. Since such a development is beyond the scope of this paper, we provide a shortcut building upon existing results under $\prod_{i=1}^n P_0$, at the cost of slowing down the rate only a little bit. To illustrate, we first note that existing results—see, for example, MacGibbon, Lu and Younes (2002), Groeneboom and Jongbloed (2012) and Durot and Lopuhaä (2014)—imply under regularity conditions that,

$$\|\mathcal{M}(-\hat{\Lambda}_n) - (-\hat{\Lambda}_n)\|_\infty = O_p((n^{-1}\log n)^{2/3}) \text{ under } \prod_{i=1}^n P_0 \ . \tag{B.15}$$

Therefore, we must have that, for any $\ell_n \uparrow \infty$ sufficiently slow,

$$\|\mathcal{M}(-\hat{\Lambda}_n) - (-\hat{\Lambda}_n)\|_\infty = o_p((n^{-1}\log n)^{2/3}\ell_n) \text{ under } \prod_{i=1}^n P_0 \ . \tag{B.16}$$

Since $\prod_{i=1}^n P_0$ and $\prod_{i=1}^n P_{\eta/\sqrt{n}}$ are mutually contiguous (see the proof of Theorem 3.2), it follows by Le Cam's first lemma that

$$\|\mathcal{M}(-\hat{\Lambda}_n) - (-\hat{\Lambda}_n)\|_\infty = o_p((n^{-1}\log n)^{2/3}\ell_n) \text{ under } \prod_{i=1}^n P_{\eta/\sqrt{n}} \ . \tag{B.17}$$

Thus, Assumption B.2(ii) is satisfied with $s_n = (n^{-1}\log n)^{2/3}\ell_n$ for a suitable $\ell_n \uparrow$. For example, if $\ell_n = (\log n)^{1/3}$, then $s_n = n^{-2/3}\log n$. $\square$

Example B.2 (cont.). In verifying Assumptions B.1(iii) and B.2(i), we focus on the pointwise asymptotics, as a through investigation of regular/efficient

estimation of Λ as in Chapter 5 in Bickel et al. (1998) is beyond the scope of this paper—and we could not find existing results in this regard. Nonetheless, we provide a sketch here. Let p_z and p_ϵ be the densities of Z_1 and ϵ_1 with respect to the Lebesgue measure. Then the joint density p_x of $X_1 \equiv (Y_1, Z_1)'$ at $x \equiv (y, z)'$ is $p_\epsilon(y - m(z))p_z(z)$. Thus, the differentiable path $\{P_{\eta/\sqrt{n}}\}$ may be constructed by perturbing the density p_ϵ, which causes no changes in Λ in view of (B.5), or perturbing m and/or p_z. Here, by "perturbing" we mean a pathwise differentiable sequence as in Definition 2.1 in the case of p_z, or, in the case of m, a scalar-parametrization $\eta \mapsto m_\eta \in \mathbf{H}$ for some suitable Hilbert space $\mathbf{H}$ such that $m_\eta = m + \eta g + o(\eta)$ as $\eta \to 0$ for some $g \in \mathbf{H}$. A thorough treatment in this regard may be found in van der Vaart (1991). Since pathwise differentiability is Hadamard differentiability—see, for example, Bickel et al. (1998, p.456), it follows by Lemmas 3.9.23 and 3.9.27 in van der Vaart and Wellner (1996) in conjunction with Lemma 3.9.3 in van der Vaart and Wellner (1996) that $\eta \mapsto \Lambda_\eta$ is pathwise differentiable where, for any $t \in [0, 1]$,

$$\Lambda_\eta(t) \equiv \int_0^t m_\eta(Q_\eta(u)) \, \mathrm{d}u \ , \tag{B.18}$$

with the quantile function Q_η arising from perturbing p_z. This completes sketching the verification of Assumption B.1(ii).

Next, we verify Assumption B.2(i) under pointwise asymptotics (i.e., $\eta = 0$); the general case may be handled with the help of Theorem 3.10.7 and Lemma 3.10.11 in van der Vaart and Wellner (1996), in a manner similar to Theorem 3.10.12 in van der Vaart and Wellner (1996). Let $\{Z_i\}_{i=1}^n$ be arranged in ascending order as $\{Z_{(i)}\}_{i=1}^n$, and let the corresponding $\{Y_i\}_{i=1}^n$ and $\{\epsilon_i\}_{i=1}^n$ be denoted by $\{Y_{(i)}\}_{i=1}^n$ and $\{\epsilon_{(i)}\}_{i=1}^n$. Then the sample cumulative sum diagram is given by $\{(k, S_n(k)) : k = 0, \ldots, n\}$ where $S_n(0) = 0$ and, for $k = 1, \ldots, n$,

$$S_n(k) = \sum_{j=1}^k Y_{(j)} \ .$$

The cumulative sum diagram plays an important role in isotonic regression because the isotonic regression estimator of m evaluated at $Z_{(i)}$ is precisely the left derivative at i of the greatest convex minorant over $[0, n]$ of $\{Y_{(i)}\}_{i=1}^n$— see Brunk (1958) and Mukerjee (1988). Following Beare and Fang (2017), we identify S_n as a random element $\hat{\Lambda}_n \in \ell^\infty([0, 1])$ defined by

$$\hat{\Lambda}_n(t) = \frac{1}{n} \sum_{i=0}^{[nt]} Y_{(i)} + \frac{nt - [nt]}{n} Y_{([nt]+1)} \tag{B.19}$$

for all $t \in [0, 1]$, where we define $Y_{(0)} = Y_{(n+1)} = 0$ and for $x \in \mathbf{R}$ we denote by $[x]$ the largest integer in $[x - 1, x]$. Proposition A.1 in Beare and Fang (2017) then shows that, under regularity conditions,

$$\sqrt{n}\{\hat{\Lambda}_n - \Lambda_{P_0}\} \xrightarrow{L} \mathbb{G} \text{ under } \prod_{i=1}^n P_0 \ , \tag{B.20}$$

for some Gaussian process $\mathbb{G} \in C([0, 1])$.

For Assumption B.3, we propose the paired bootstrap, i.e., resample from $\{Y_i, Z_i\}_{i=1}^n$ with replacement, and construct $\hat{\Lambda}_n^*$ in the same fashion as $\hat{\Lambda}_n$. The main arguments underlying Proposition A.1 in Beare and Fang (2017) are the Delta method and weak convergence of the partial sum defined by the errors. Thus, the bootstrap consistency of $\hat{\mathbb{G}}_n \equiv \sqrt{n}\{\hat{\Lambda}_n^* - \hat{\Lambda}_n\}$ follows from Theorem 3.9.4 in van der Vaart and Wellner (1996), and consistency of bootstrapping the partial sum process for which we refer the reader to Kinateder (1992), Holmes, Kojadinovic and Quessy (2013) and Calhoun (2018) for related results. □

References

ALIPRANTIS, C. D. and BORDER, K. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Third ed. Springer-Verlag, Berlin. MR2378491

ANDREWS, D. W. K. and GUGGENBERGER, P. (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory* **26** 426–468. MR2600570

BALABDAOUI, F. and WELLNER, J. A. (2007). A Kiefer-Wolfowitz theorem for convex densities. In *Asymptotics: Particles, Processes and Inverse Problems. Lecture Notes–Monograph Series* **55** 1–31. Institute of Mathematical Statistics. MR2435342

BANERJEE, M. (2005). Likelihood ratio tests under local and fixed alternatives in monotone function problems. *Scandinavian Journal of Statistics* **32** 507–525. MR2232340

BANERJEE, M. and SAMWORTH, R. J. (2018). A conversation with Jon Wellner. *Statistical Science* **33** 633–651. MR3881212

BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *The Annals of Statistics* **29** 1699–1731. MR1891743

BEARE, B. K. and FANG, Z. (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electronic Journal of Statistics* **11** 3841–3870. MR3714300

BEARE, B. K. and MOON, J.-M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory* **31** 471–492. MR3348455

BEARE, B. and SHI, X. (2019). An improved bootstrap test of density ratio ordering. *Econometrics and Statistics* **10** 9–26. MR3945178

BICKEL, P. J., KLAASSEN, C. A., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, New York. MR1623559

BRUNK, H. D. (1958). On the estimation of parameters restricted by inequalities. *The Annals of Mathematical Statistics* **29** 437–454. MR0132632

BURCHARD, H. G. (1974). Splines (with optimal knots) are better. *Applicable Analysis* **3** 309–319. MR0399708

CALHOUN, G. (2018). Block bootstrap consistency under weak assumptions. *Econometric Theory* **34** 1383–1406. MR3873853

CAROLAN, C. A. and TEBBS, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two-sample problem. *Biometrika* **92** 159–171. MR2158617

Cheng, G. and Huang, J. (2010). Bootstrap consistency for general semiparametric M-estimation. *The Annals of Statistics* **38** 2884–2915. MR2722459

Chetverikov, D. (2019). Testing regression monotonicity in econometric models. *Econometric Theory* **35** 729–776. MR3974977

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society. MR1604537

de Boor, C. (1973). Good approximation by splines with variable knots. In *Spline Functions and Approximation Theory. ISNM International Series of Numerical Mathematics* **21** 57–72. Birkhäuser Basel. MR0403169

de Boor, C. (2001). *A Practical Guide to Splines*, Revised ed. Springer-Verlag. MR1900298

Delgado, M. A. and Escanciano, J. C. (2012). Distribution-free tests of stochastic monotonicity. *Journal of Econometrics* **170** 68–75. MR2955940

Delgado, M. A. and Escanciano, J. C. (2013). Conditional stochastic dominance testing. *Journal of Business & Economic Statistics* **31** 16–28. MR3030795

Delgado, M. A. and Escanciano, J. C. (2016). Distribution-free tests of conditional moment inequalities. *Journal of Statistical Planning and Inference* **173** 99–108. MR3465015

Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields* **95** 125–140. MR1207311

Durot, C. (2003). A Kolmogorov-type test for monotonicity of regression. *Statistics & Probability Letters* **63** 425–433. MR1996191

Durot, C. and Lopuhaä, H. P. (2014). A Kiefer-Wolfowitz type of result in a general setting, with an application to smooth monotone estimation. *Electronic Journal of Statistics* **8** 2479–2513. MR3285873

Durot, C. and Tocquet, A.-S. (2003). On the distance between the empirical process and its concave majorant in a monotone regression framework. *Annales de l'Institut Henri Poincare (B) Probability and Statistics* **39** 217–240. MR1962134

Fang, Z. (2019). Refinements of the Kiefer-Wolfowitz Theorem and a Test of Concavity. arXiv preprint: arXiv:1906.10305.

Fang, Z. and Santos, A. (2015). Inference on directionally differentiable functions. arXiv preprint: arXiv:1404.3763. MR3936869

Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies* **86** 377–412. MR3936869

Giné, E. t. and Zinn, J. (1990). Bootstrapping general empirical measures. *The Annals of Probability* **18** 851–869. MR1055437

Grenander, U. (1956). On the theory of mortality measurement: Part II. *Scandinavian Actuarial Journal* **1956** 125–153. MR0093415

Groeneboom, P. and Jongbloed, G. (2012). Isotonic L_2-projection test for local monotonicity of a hazard. *Journal of Statistical Planning and Inference* **142** 1644–1658. MR2903377

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints*. Cambridge University Press. MR3445293

HOLMES, M., KOJADINOVIC, I. and QUESSY, J.-F. (2013). Nonparametric tests for change-point detection à la Gombay and Horváth. *Journal of Multivariate Analysis* **115** 16–32. MR3004542

KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **34** 73–85. MR0397974

KIEFER, J. and WOLFOWITZ, J. (1977). Asymptotically minimax estimation of concave and convex distribution functions II. In *Statistical Decision Theory and Related Topics II* 193–211. Academic Press. MR0443202

KINATEDER, J. G. (1992). An invariance principle applicable to the bootstrap. In *Exploring the Limits of Bootstrap*, **270** 157–181. John Wiley & Sons, New York. MR1197784

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York. MR2724368

KULIKOV, V. N. and LOPUHAÄ, H. P. (2004). Testing for a Monotone Density Using L_k-Distances between the Empirical Distribution Function and Its Concave Majorant. Technical report: https://www.eurandom.tue.nl/reports/2004/028-report.pdf.

KULIKOV, V. N. and LOPUHAÄ, H. P. (2008). Distribution of global measures of deviation between the empirical distribution function and its concave majorant. *Journal of Theoretical Probability* **21** 356–377. MR2391249

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3 ed. Springer, New York. MR2135927

MACGIBBON, B., LU, J. and YOUNES, H. (2002). Limit theorems for asymptotically minimax estimation of a distribution with increasing failure rate under a random mixed censorship/truncation model. *Communications in Statistics-Theory and Methods* **31** 1309–1333. MR1920851

MUKERJEE, H. (1988). Monotone nonparametric regression. *The Annals of Statistics* **16** 741–750. MR0947574

PFANZAGL, J. (1985). *Asymptotic Expansions for General Statistical Models.* Springer, Berlin (with the assistance of Wefelmeyer W). MR0810004

PRAESTGAARD, J. and WELLNER, J. (1993). Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability* **21** 2053–2086. MR1245301

ROMANO, J. P. and SHAIKH, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics* **40** 2798–2822. MR3097960

SEO, J. (2018). Tests of stochastic monotonicity with improved power. *Journal of Econometrics* **207** 53–70. MR3856760

SHAPIRO, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications* **66** 477–487. MR1080259

SHAPIRO, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research* **30** 169–186. MR1118896

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons. MR0838963

VAN DER VAART, A. W. (1991). On differentiable functionals. *The Annals of*

Statistics **19** 178–204. MR1091845

VAN DER VAART, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK MR1652247

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York. MR1385671

WOODROOFE, M. and SUN, J. (1999). Testing uniformity versus a monotone density. *The Annals of Statistics* **27** 338–360. MR1701114

ZAANEN, A. C. (1997). *Introduction to Operator Theory in Riesz Spaces.* Springer. MR1631533