Electronic Journal of Statistics
Vol. 11 (2017) 3841–3870
ISSN: 1935-7524
DOI: 10.1214/17-EJS1349

Weak convergence of the least concave majorant of estimators for a concave distribution function*

Brendan K. Beare

Department of Economics
University of California, San Diego
La Jolla, CA 92093
e-mail: bbeare@ucsd.edu

and

Zheng Fang

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

Abstract: We study the asymptotic behavior of the least concave majorant of an estimator of a concave distribution function under general conditions. The true concave distribution function is permitted to violate strict concavity, so that the empirical distribution function and its least concave majorant are not asymptotically equivalent. Our results are proved by demonstrating the Hadamard directional differentiability of the least concave majorant operator. Standard approaches to bootstrapping fail to deliver valid inference when the true distribution function is not strictly concave. While the rescaled bootstrap of Dümbgen delivers asymptotically valid inference, its performance in small samples can be poor, and depends upon the selection of a tuning parameter. We show that two alternative bootstrap procedures—one obtained by approximating a conservative upper bound, the other by resampling from the Grenander estimator—can be used to construct reliable confidence bands for the true distribution. Some related results on isotonic regression are provided.

MSC 2010 subject classifications: Primary 62G09, 62G20.
Keywords and phrases: Least concave majorant, Hadamard directional derivative, Grenander estimator, rescaled bootstrap.

Received October 2016.

1. Introduction

Nonparametric estimation under shape constraints such as monotonicity and concavity has received increasing attention in recent years. Groeneboom and

*We thank Kengo Kato, Andres Santos, Stanislav Volgushev, seminar participants at the University of Tokyo, and the anonymous referees for helpful comments.

Jongbloed (2014) provide a helpful introduction to the current state of the field. As pointed out by Walther (2009), nonparametric estimation under shape constraints is attractive for two main reasons: (1) shape constraints are often implied by theoretical models or are at least plausible assumptions, and (2) nonparametric estimation under shape constraints is often feasible without the use of tuning parameters, as opposed to classical kernel or series estimators. Perhaps the best known shape constrained estimator is the Grenander estimator of a nonincreasing density function. Grenander (1956) showed that, given a random sample drawn from a nonincreasing probability density, the left-derivative of the least concave majorant (LCM) of the empirical distribution function achieves the maximum likelihood among all nonincreasing densities.

In this paper we provide some new results concerning the asymptotic behavior of the least concave majorant of an empirical distribution function. Quite a lot is known about its left-derivative, the Grenander estimator, already. Denote by $\hat{f}_n$ the Grenander estimator of a nonincreasing density f based on a sample of size n with empirical distribution $\mathbb{F}_n$, and by $\hat{\mathbb{F}}_n$ the LCM of $\mathbb{F}_n$. The pointwise asymptotic distribution of $\hat{f}_n$ was obtained by Prakasa Rao (1969) at points where f is strictly decreasing, and by Carolan and Dykstra (1999) and Jankowski (2014) at points where f is flat or misspecified. The rate of convergence is $n^{1/3}$ in the former case and $n^{1/2}$ in the latter, and the limit distribution is non-Gaussian. Results on the global asymptotic behavior of $\hat{f}_n$—specifically, of its L_p-risk—have been provided by Groeneboom (1985), Groeneboom et al. (1999), Kulikov and Lopuhaä (2005), Durot (2007) and Durot et al. (2012). These global results require f to be strictly decreasing on its support. Kulikov and Lopuhaä (2006a) studied the behavior of $\hat{f}_n$ near the boundary of its support, while Woodroofe and Sun (1993) and Balabdaoui et al. (2011) studied its behavior at zero.

Turning to the asymptotic behavior of $\hat{\mathbb{F}}_n$, it is natural to consider weak convergence of the process $\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{F}}_n - F)$. A result of Kiefer and Wolfowitz (1976) implies that $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ are asymptotically equivalent when F satisfies strict concavity on its support, so that $\hat{\mathbb{G}}_n$ converges weakly to $\mathbb{G} = \mathbb{B} \circ F$, a Brownian bridge $\mathbb{B}$ composed with F. Other results on the behavior of $\hat{\mathbb{F}}_n - \mathbb{F}_n$ when F is strictly concave on its support have been provided by Wang (1994) and Kulikov and Lopuhaä (2006b, 2008). On the other hand, when F is the uniform distribution on the unit interval it is known that $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ are not asymptotically equivalent, and $\hat{\mathbb{G}}_n$ converges weakly to the LCM of $\mathbb{B}$, a process studied in detail by Carolan and Dykstra (2001). Carolan (2002) considered the more general case where F is affine over some maximal subinterval $[a, b]$ of its support, and showed that the restriction of $\hat{\mathbb{G}}_n$ to $[a, b]$ converges weakly to the LCM of the restriction of $\mathbb{G}$ to $[a, b]$.

Our main result, Theorem 2.1 below, establishes the weak convergence of $\hat{\mathbb{G}}_n$ in the intermediate cases where F is concave but not necessarily strictly concave or uniform. As might be guessed from the results of Carolan (2002), the weak limit $\hat{\mathbb{G}}$ can be obtained by taking LCMs of $\mathbb{G}$ over the distinct intervals on which F is affine. Our proof exploits the fact that the LCM operator is Hadamard directionally differentiable (see Definition 2.2 and Proposition 2.1

below) despite not being fully Hadamard differentiable. This provides enough structure to invoke the Delta method (Shapiro, 1991; Dümbgen, 1993) and in this way derive the weak limit of $\hat{\mathbb{G}}_n$. Our result applies not only to the LCM of an empirical distribution function, but to any estimator of F obtained by taking the LCM of an estimator $\mathbb{F}_n$ of F for which $\sqrt{n}(\mathbb{F}_n - F)$ converges weakly to a continuous process $\mathbb{G}$ vanishing at infinity. Thus, for instance, we may take $\mathbb{F}_n$ to be a smoothed estimate of F, as in Eggermont and LaRiccia (2000), and we may allow the data used to construct $\mathbb{F}_n$ to exhibit limited serial dependence. We allow F to have bounded or unbounded support on the nonnegative half-line.

Our weak convergence result for $\hat{\mathbb{G}}_n$ can be useful, for example, when constructing uniform confidence bands for F. However, doing so requires consistent estimation of the law of the weak limit $\hat{\mathbb{G}}$, which turns out to be nontrivial. The fact that the LCM operator fails to be fully Hadamard differentiable implies that the bootstrap does not produce consistent estimates of the law of $\hat{\mathbb{G}}$ (Dümbgen, 1993; Fang and Santos, 2015). In other words, the Delta method generalizes under Hadamard directional differentiability to obtaining the weak limit of $\hat{\mathbb{G}}_n$ but not to obtaining bootstrap consistency. We show in Theorem 3.1 that it instead approximates a process that coincides with $\hat{\mathbb{G}}$ only when F is strictly concave. The rescaled bootstrap of Dümbgen (1993) does produce consistent estimates of the law of $\hat{\mathbb{G}}$ (Theorem 3.2) but, consistent with Dümbgen's warning of nonrobustness, we find that it performs poorly in numerical simulations. We suggest two alternative bootstrap procedures that can be used to construct reliable confidence bands for a concave distribution function. One is obtained by approximating the distribution of the uniform norm of $\mathbb{G}$, which stochastically dominates the uniform norm of $\hat{\mathbb{G}}$ and can be consistently bootstrapped. The other is obtained by resampling from the Grenander estimator. This latter procedure is interesting in that, while it fails to provide consistent estimates of the law of $\hat{\mathbb{G}}$, it does provide unconditionally consistent estimates of upper quantiles of the law of the uniform norm of $\hat{\mathbb{G}}$. Theorem 4.1 identifies the process approximated by bootstrapping from the Grenander estimator, and numerical simulations are used to show that the upper quantiles of its uniform norm coincide with the upper quantiles of the uniform norm of $\hat{\mathbb{G}}$.

In Appendix A we briefly discuss how our approach to studying the asymptotic behavior of estimators of concave distribution functions may be extended to the study of isotonic regression. Specifically, we develop limit theory for the greatest convex minorant of the cumulative sum diagram, whose left-derivative is the isotonic regression estimator. Weak convergence of a suitably normalized version of the cumulative sum diagram is established in Proposition A.1, and of its greatest convex minorant in Proposition A.2.

Before proceeding further we introduce some additional notation. We denote by $\mathbf{R}^+$ the set $\{x \in \mathbf{R} : x \geq 0\}$. The underlying probability space on which all random elements are defined is $(\Omega, \mathcal{F}, P)$. For a set T, we let $\ell^\infty(T)$ denote the set of uniformly bounded, real valued functions on T. Of particular importance is the space $\ell^\infty = \ell^\infty(\mathbf{R}^+)$, which we equip with the uniform metric d^∞ and ball σ-field $\mathcal{A}$. Random elements of ℓ^∞ are $\mathcal{F}/\mathcal{A}$-measurable maps from Ω to ℓ^∞. We

denote by $\rightsquigarrow$ weak convergence in (ℓ^∞, d^∞) in the sense of Hoffmann-Jørgensen. We denote by $C_0 = C_0(\mathbf{R}^+)$ the collection of continuous real valued functions on $\mathbf{R}^+$ vanishing at infinity, also equipped with d^∞. We denote by $\|\cdot\|$ the uniform norm on $\ell^\infty(T)$, where T should be clear from context.

2. Weak convergence of $\hat{\mathbb{G}}_n$ to $\hat{\mathbb{G}}$

The distribution function F to be estimated is taken to be that of a nonnegative real valued random variable. We treat it as an element of ℓ^∞ and maintain throughout that it satisfies the following condition.

Assumption 2.1. $F : \mathbf{R}^+ \to \mathbf{R}$ is a concave distribution function.

No further technical conditions will be directly imposed on F. To maintain generality about the underlying sampling scheme and method of estimation, we suppose the existence of a sequence $\mathbb{F}_1, \mathbb{F}_2, \ldots$ of random elements of ℓ^∞ satisfying the following high level condition, in which $\mathbb{G}_n$ denotes $\sqrt{n}(\mathbb{F}_n - F)$.

Assumption 2.2. $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ for some random element $\mathbb{G}$ of ℓ^∞ with sample paths in C_0.

If $\mathbb{F}_n$ is the empirical distribution function of an independent and identically distributed (iid) sample of size n drawn from F then $\mathbb{G}_n$ is the usual empirical process and clearly Assumption 2.2 is satisfied with $\mathbb{G} = \mathbb{B} \circ F$ and $\mathbb{B}$ a Brownian bridge. More generally, we may allow the sample drawn from F to satisfy a mixing condition or related property (Dehling and Philipp, 2002). The ball σ-field $\mathcal{A}$ on ℓ^∞ is coarse enough to accommodate empirical distribution functions as random elements of ℓ^∞, unlike the Borel σ-field on ℓ^∞ (cf. Pollard, 1984, pp. 65-66). We might also take $\mathbb{F}_n$ to be a smoothed empirical distribution function (van der Vaart, 1994) or allow $\mathbb{F}_n$ to be some other estimator satisfying Assumption 2.2 under suitable regularity conditions.

To exploit the concavity of F we propose using the estimator $\hat{\mathbb{F}}_n = \mathcal{M}\mathbb{F}_n$, where $\mathcal{M}$ is the LCM operator. If $\mathbb{F}_n$ is the empirical distribution function of an iid sample drawn from F, then the left-derivative of $\hat{\mathbb{F}}_n$ is the classical Grenander estimator of the probability density for F. The following definition is adapted from Beare and Moon (2015, Def. 2.1).

Definition 2.1. Given a nonempty convex set $T \subseteq \mathbf{R}^+$, the LCM over T is the operator $\mathcal{M}_T : \ell^\infty \to \ell^\infty(T)$ that maps each $\theta \in \ell^\infty$ to the function

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

Note that $\theta(x) \leq \mathcal{M}_T\theta(x) \leq \sup_{x' \in T} \theta(x')$ for all $x \in T$, so that we do in fact have $\mathcal{M}_T\theta \in \ell^\infty(T)$ for each $\theta \in \ell^\infty$. We write $\mathcal{M}$ as shorthand for $\mathcal{M}_{\mathbf{R}^+}$ and refer to $\mathcal{M}$ as the LCM operator.

The definition of $\mathcal{M}_T$ given here differs from that of Beare and Moon (2015) only in that those authors took the domain of $\mathcal{M}_T$ to be $\ell^\infty([0,1])$ and required T to be a closed subinterval of the unit interval. It is easy to show that $\mathcal{M}_T\theta$

is a concave majorant of θ by arguing as in Eggermont and LaRiccia (2001, pp. 225–226). Other well known or easily established properties of $\mathcal{M}_T$ include monotonicity, convexity, positive homogeneity of degree one, and weak contractivity (Durot and Tocquet, 2003, Lem. 2.2): $\|\mathcal{M}_T\theta_1 - \mathcal{M}_T\theta_2\| \leq \|\theta_1 - \theta_2\|$. Beare and Moon (2015) investigated the differential properties of $\mathcal{M}_{[0,1]}$ in order to study the asymptotic behavior of a test of the monotone density ratio property proposed by Carolan and Tebbs (2005). For an application of this test to an empirical puzzle in the financial literature, see Beare and Schmidt (2016). Related tests of shape constraints based on the LCM operator have been developed by Delgado and Escanciano (2012, 2013, 2016).

As pointed out by Beare and Moon (2015), $\mathcal{M}_{[0,1]}$ fails to be Hadamard differentiable and hence violates the assumptions made in standard treatments of the Delta method (see e.g. van der Vaart and Wellner, 1996, Thm. 3.9.4); however, $\mathcal{M}_{[0,1]}$ does satisfy a certain form of directional differentiability introduced by Shapiro (1990). Quite remarkably, the Delta method is valid under this weaker notion of differentiability (Shapiro, 1991), although the Delta method for the bootstrap typically fails (Dümbgen, 1993). Fang and Santos (2015) discuss the contributions of Shapiro and Dümbgen at length, providing a range of extensions, and illustrating their applicability to a variety of problems in econometrics.

Definition 2.2. Let $\mathbf{D}$ and $\mathbf{E}$ be Banach spaces. A map $\phi : \mathbf{D} \to \mathbf{E}$ is said to be *Gâteaux directionally differentiable* at $\theta \in \mathbf{D}$ tangentially to a set $\mathbf{D}_0 \subset \mathbf{D}$ if there is a map $\phi'_\theta : \mathbf{D}_0 \to \mathbf{E}$ such that

$$\left\| \frac{\phi(\theta + t_n h) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathbf{E}} \to 0 \tag{2.1}$$

for all $h \in \mathbf{D}_0$ and all $t_1, t_2, \ldots \in \mathbf{R}^+$ such that $t_n \downarrow 0$. It is said to be *Hadamard directionally differentiable* at $\theta \in \mathbf{D}$ tangentially to a set $\mathbf{D}_0 \subset \mathbf{D}$ if there is a map $\phi'_\theta : \mathbf{D}_0 \to \mathbf{E}$ such that

$$\left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathbf{E}} \to 0 \tag{2.2}$$

for all $h \in \mathbf{D}_0$ and all $h_1, h_2, \ldots \in \mathbf{D}$ and $t_1, t_2, \ldots \in \mathbf{R}^+$ such that $t_n \downarrow 0$ and $\|h_n - h\|_{\mathbf{D}} \to 0$.

As with various notions of differentiability in the literature, Hadamard directional differentiability can be understood by looking at the restrictions imposed on the approximating map (i.e. the derivative) and the way the approximation error is controlled. Specifically, define the remainder term

$$\mathrm{Rem}_\theta(h) \equiv \phi(\theta + h) - (\phi(\theta) + \phi'_\theta(h)), \tag{2.3}$$

where $\phi(\theta) + \phi'_\theta(h)$ can be viewed as the first order approximation to $\phi(\theta + h)$. Hadamard directional differentiability of ϕ implies that the approximation error

$\mathrm{Rem}_\theta(h)$ satisfy that $\mathrm{Rem}_\theta(th)/t$ tends to zero uniformly in $h \in K$ for any compact set K – i.e.,

$$\sup_{h\in K} \left\| \frac{\mathrm{Rem}_\theta(th)}{t} \right\|_{\mathbf{E}} \to 0 \quad \text{as} \quad t \downarrow 0. \tag{2.4}$$

Compact differentiability in this sense, together with continuity of ϕ'_θ, is equivalent to Hadamard directional differentiability (Shapiro, 1990, Prop. 3.3). However, while continuity of ϕ'_θ is assured under Hadamard directional differentiability, linearity is often lost. Gâteaux directional differentiability is of course weaker than Hadamard directional differentiability, as no robustness with respect to perturbations in the direction h is required.

Our first result, similar to Lemma 3.2 of Beare and Moon (2015), establishes Hadamard directional differentiability of the LCM operator. Its proof, along with the proofs of other results to be stated, may be found in Section 5 below.

Proposition 2.1. *The LCM operator $\mathcal{M} : \ell^\infty \to \ell^\infty$ is Hadamard directionally differentiable at any concave $\theta \in \ell^\infty$ tangentially to C_0. Its directional derivative $\mathcal{M}'_\theta : C_0 \to \ell^\infty$ is uniquely determined as follows: for any $h \in C_0$ and $x \in \mathbf{R}^+$, we have $\mathcal{M}'_\theta h(x) = \mathcal{M}_{T_{\theta,x}} h(x)$, where $T_{\theta,x} = \{x\} \cup U_{\theta,x}$, and $U_{\theta,x}$ is the union of all open intervals $A \subset \mathbf{R}^+$ such that (1) $x \in A$, and (2) θ is affine on A.*

Remark 2.1. A concave $\theta \in \ell^\infty$ is nondecreasing, and is continuous on $(0, \infty)$.

Remark 2.2. If θ is affine on an open interval containing $x \in (0, \infty)$, then the set $T_{\theta,x}$ is the maximal such open interval. If θ is not affine on an open interval containing $x \in (0, \infty)$, or if $x = 0$, then we have $T_{\theta,x} = \{x\}$ and $\mathcal{M}'_\theta h(x) = h(x)$. The directional derivative $\mathcal{M}'_\theta$ therefore behaves like a hybrid of the LCM and identity operators: for any direction $h \in C_0$, it majorizes h by concave functions on regions over which θ is affine but acts like an identity map elsewhere.

Our next result summarizes some useful properties of $\mathcal{M}'_\theta$.

Proposition 2.2. *For each $\theta \in \ell^\infty$, the Hadamard directional derivative $\mathcal{M}'_\theta$ is weakly contractive and convex. Further, $\mathcal{M}'_\theta$ is linear if and only if θ is strictly concave on $\mathbf{R}^+$, in which case $\mathcal{M}'_\theta h = h$ for all $h \in C_0$.*

Let $\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{F}}_n - F)$ and let $\hat{\mathbb{G}} = \mathcal{M}'_F \mathbb{G}$. With Hadamard directional differentiability of $\mathcal{M}$ in hand, we obtain the weak limit of $\hat{\mathbb{G}}_n$ by employing the Delta method.

Theorem 2.1. *Under Assumptions 2.1 and 2.2 we have $\hat{\mathbb{G}}_n \rightsquigarrow \hat{\mathbb{G}}$.*

Remark 2.3. Theorem 2.1 can be viewed as an extension of a result of Carolan (2002), who showed in the proof of his Theorem 5 that when F is affine over a maximal interval $[a, b] \subseteq \mathbf{R}^+$, the restriction of $\hat{\mathbb{G}}_n$ to $[a, b]$ converges weakly to the LCM of the restriction of $\mathbb{G}$ to $[a, b]$. Our result extends his to obtain weak convergence of the entire process $\hat{\mathbb{G}}_n$ even when F may have multiple affine segments separated by kinks, or by intervals over which it is strictly concave. Further, since our proof is an application of the Delta method, it is simple for us

to consider general estimators $\mathbb{F}_n$ satisfying Assumption 2.2, whereas Carolan (2002) requires $\mathbb{F}_n$ to be the empirical distribution based on iid draws from F.

Remark 2.4. For typical choices of $\mathbb{F}_n$ satisfying Assumption 2.2 we will have $\hat{\mathbb{G}}(x) = 0$ for all x such that $F(x) = 1$.

Remark 2.5. If F is strictly concave then we know from Proposition 2.2 that $\mathcal{M}'_F$ is the identity map on C_0. In this case $\hat{\mathbb{G}} = \mathbb{G}$, and so Theorem 2.1 implies the asymptotic equivalence of $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$. More generally, asymptotic equivalence of $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ holds when F is strictly concave at all x such that $F(x) < 1$, and $\mathbb{G}(x) = 0$ for all x such that $F(x) = 1$

3. Bootstrap approximation of the law of $\hat{\mathbb{G}}_n$

Parallel to the level of generality adopted in our treatment of the estimator $\mathbb{F}_n$ in the previous section, we here maintain a high degree of generality with respect to the method used to obtain a bootstrap version of $\mathbb{F}_n$. Let $\mathbb{F}_1^*, \mathbb{F}_2^*, \ldots$ be random elements of ℓ^∞. For each n we assume the existence of independent σ-fields $\mathcal{X}_n, \mathcal{W}_n^* \subset \mathcal{F}$ such that $\mathbb{F}_n$ is $\mathcal{X}_n/\mathcal{A}$-measurable and $\mathbb{F}_n^*$ is $(\mathcal{X}_n \otimes \mathcal{W}_n^*)/\mathcal{A}$-measurable. $\mathcal{X}_n$ should be interpreted as the σ-field generated by the data used to construct $\mathbb{F}_n$, and $\mathcal{W}_n^*$ should be interpreted as the σ-field generated by the random weights associated with the bootstrap procedure used to generate a realization of $\mathbb{F}_n^*$. For instance, if $\mathbb{F}_n$ is the empirical distribution of an iid sample $(X_1, \ldots, X_n)$ drawn from F, then the standard nonparametric bootstrap involves computing

$$\mathbb{F}_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} W_{i,n}^* 1(X_i \le x), \quad x \in \mathbf{R}^+,$$

where $(W_{1,n}^*, \ldots, W_{n,n}^*)$ is a multinomial random vector with n categories and probabilities $(n^{-1}, \ldots, n^{-1})$, distributed independent of $(X_1, \ldots, X_n)$. In this case $\mathcal{X}_n$ is the σ-field generated by $(X_1, \ldots, X_n)$ and $\mathcal{W}_n^*$ is the σ-field generated by $(W_{1,n}^*, \ldots, W_{n,n}^*)$. More general resampling schemes such as block bootstraps or Markovian procedures for serially dependent data may also be accommodated within our framework; see Radulović (2002) for a survey and discussion.

We require some additional notation. Let $\mathcal{P}$ denote the collection of all probability measures on the measurable space $(\ell^\infty, \mathcal{A})$. Denote by $d_{\mathcal{P}}$ a metric on $\mathcal{P}$ that metrizes weak convergence in ℓ^∞ under d_∞; for instance, we could take $d_{\mathcal{P}}$ to be the bounded Lipschitz metric, as in Pollard (1984, pp. 74–75). Given an arbitrary random element $\mathbb{H}$ of ℓ^∞ and σ-field $\mathcal{C} \subseteq \mathcal{F}$, denote by $\mathcal{L}(\mathbb{H})$ the law of $\mathbb{H}$ and by $\mathcal{L}(\mathbb{H} \,|\, \mathcal{C})$ the law of $\mathbb{H}$ conditional on $\mathcal{C}$, the latter uniquely defined up to almost sure equivalence.

Let $\mathbb{G}_n^* = \sqrt{n}(\mathbb{F}_n^* - \mathbb{F}_n)$, the bootstrap version of $\mathbb{G}_n$. We assume that the bootstrap law of $\mathbb{G}_n^*$ delivers a valid approximation to the law of $\mathbb{G}_n$, in the following sense.

Assumption 3.1. $\mathbb{G}_n^*$ satisfies $d_{\mathcal{P}}\left(\mathcal{L}(\mathbb{G}_n^* \,|\, \mathcal{X}_n), \mathcal{L}(\mathbb{G}_n)\right) \to 0$ in outer probability.

Assumption 3.1 is automatically satisfied when $\mathbb{F}_n^*$ is the standard nonparametric bootstrap version of the empirical distribution of iid draws from F. It is also satisfied by various dependent resampling schemes under suitable technical conditions, as discussed by Radulović (2002).

We would like to obtain a valid bootstrap approximation to the law of $\hat{\mathbb{G}}_n$. An obvious approach is to use as an approximation the law of $\sqrt{n}(\mathcal{M}\mathbb{F}_n^* - \mathcal{M}\mathbb{F}_n)$ conditional on $\mathcal{X}_n$. If we could apply the Delta method for the bootstrap (see e.g. van der Vaart and Wellner, 1996, Thm. 3.9.11) it would be possible to conclude that

$$d_{\mathcal{P}}(\mathcal{L}\left(\sqrt{n}(\mathcal{M}\mathbb{F}_n^* - \mathcal{M}\mathbb{F}_n) \,|\, \mathcal{X}_n\right), \mathcal{L}(\hat{\mathbb{G}})) \to 0 \tag{3.1}$$

in outer probability, justifying the use of the law of $\sqrt{n}(\mathcal{M}\mathbb{F}_n^* - \mathcal{M}\mathbb{F}_n)$ conditional on $\mathcal{X}_n$ as an approximation to the law of $\hat{\mathbb{G}}_n$. Unfortunately it is not possible to apply the Delta method for the bootstrap in this way when the directional derivative $\mathcal{M}'_F$ is nonlinear, and the above statement of convergence in outer probability is typically false unless F is strictly concave on its support. This is a consequence of Proposition 1 of Dümbgen (1993) and Theorem A.1 of Fang and Santos (2015); see also Volgushev and Shao (2014, p. 417). We instead have the following result characterizing the unconditional weak limit of $\sqrt{n}(\mathcal{M}\mathbb{F}_n^* - \mathcal{M}\mathbb{F}_n)$.

Theorem 3.1. *Under Assumptions 2.1, 2.2 and 3.1, we have*

$$\sqrt{n}(\mathcal{M}\mathbb{F}_n^* - \hat{\mathbb{F}}_n) \rightsquigarrow \mathcal{M}'_F(\mathbb{G} + \mathbb{G}') - \mathcal{M}'_F(\mathbb{G}'),$$

where $\mathbb{G}'$ is an independent copy of $\mathbb{G}$.

The unconditional weak limit appearing in Theorem 3.1 is equal to $\hat{\mathbb{G}}$ if F is strictly concave on its support, but otherwise typically differs from $\hat{\mathbb{G}}$. In the latter case we cannot have weak convergence of $\sqrt{n}(\mathcal{M}\mathbb{F}_n^* - \mathcal{M}\mathbb{F}_n)$ conditional on $\mathcal{X}_n$ to $\hat{\mathbb{G}}$, so the bootstrap is not well-behaved.

In view of the failure of the standard bootstrap approximation to the law of $\hat{\mathbb{G}}_n$, we may instead consider an alternative route based on a suitable estimator $\hat{\mathcal{M}}'_n$ of the operator $\mathcal{M}'_F$. The simple plug-in estimator $\mathcal{M}'_{\mathbb{F}_n}$ is not effective because from Proposition 2.1 it is clear that $\mathcal{M}'_F h$ is typically not continuous in F for fixed h. We may instead construct $\hat{\mathcal{M}}'_n$ using a version of numerical differentiation. Specifically, let $\hat{\mathcal{M}}'_n$ be given by

$$\hat{\mathcal{M}}'_n h = \frac{\mathcal{M}(\mathbb{F}_n + t_n h) - \mathcal{M}(\mathbb{F}_n)}{t_n}, \quad h \in \ell^\infty, \tag{3.2}$$

with $t_1, t_2, \ldots$ a sequence of positive reals chosen to satisfy the following condition.

Assumption 3.2. As $n \to \infty$ we have $t_n \to 0$ and $\sqrt{n}t_n \to \infty$.

A bootstrap approximation to the law of $\hat{\mathbb{G}}_n$ is provided by the law of $\hat{\mathcal{M}}'_n \mathbb{G}_n^*$ conditional on $\mathcal{X}_n$. Note that, in view of the $\mathcal{X}_n/\mathcal{A}$-measurability of $\mathbb{F}_n$ and $(\mathcal{X}_n \otimes \mathcal{W}_n^*)/\mathcal{A}$-measurability of $\mathbb{G}_n^*$, we may in practice simulate the law of

$\hat{\mathcal{M}}'_n \mathbb{G}^*_n$ conditional on $\mathcal{X}_n$ by evaluating $\hat{\mathcal{M}}'_n$ at repeated bootstrap realizations of $\mathbb{G}^*_n$. The next result indicates that this modified bootstrap procedure leads to asymptotically valid approximation of the law of $\hat{\mathbb{G}}_n$.

Theorem 3.2. *If Assumptions 2.1, 2.2, 3.1 and 3.2 hold, then*

$$d_{\mathcal{P}}(\mathcal{L}(\hat{\mathcal{M}}'_n \mathbb{G}^*_n \mid \mathcal{X}_n), \mathcal{L}(\hat{\mathbb{G}})) \to 0$$

in outer probability.

The bootstrap procedure just described is precisely the rescaled bootstrap studied by Dümbgen (1993). It can also be viewed as a special case of the more general bootstrap procedure studied by Fang and Santos (2015). Though Theorem 3.2 is immediate from Proposition 2 of Dümbgen (1993) and Proposition 2.1 above, to provide additional clarity we give a proof based on verifying the conditions of Theorem 3.2 of Fang and Santos (2015). Despite the asymptotic approximation established in Theorem 3.2, we will see in the following section that confidence bands constructed using the rescaled bootstrap perform poorly in numerical simulations.

We close this section by pointing out a connection between the rescaled bootstrap and the more familiar m-out-of-n bootstrap. This connection was identified by Dümbgen (1993), who pointed to earlier uses of the m-out-of-n bootstrap by Bretagnolle (1983) and Beran and Srivastava (1985, 1987); see also Bickel et al. (1997). Suppose we let $\mathbb{F}^*_n$ be the empirical distribution of an iid sample of $m = m(n)$ draws from $\mathbb{F}_n$, and set $\mathbb{G}^*_n = \sqrt{m}(\mathbb{F}^*_n - \mathbb{F}_n)$ and $t_n = m^{-1/2}$. Assumption 3.1 will be satisfied if, for instance, $\mathbb{F}_n$ is the empirical distribution of an iid sample from F and m tends to infinity with n (see e.g. Shorack and Wellner, 1986, pp. 763–764) while Assumption 3.2 will be satisfied if m increases to infinity more slowly than n. For this choice of $\mathbb{G}^*_n$ and t_n we have

$$\hat{\mathcal{M}}'_n \mathbb{G}^*_n = \sqrt{m}\,(\mathcal{M}(\mathbb{F}^*_n) - \mathcal{M}(\mathbb{F}_n))\,,$$

and so our rescaled bootstrap approximation to the law of $\hat{\mathbb{G}}_n$ is precisely the usual m-out-of-n bootstrap approximation. It is thus apparent that, when the m-out-of-n bootstrap works, it does so because it implicitly provides an estimate of $\mathcal{M}'_F \mathbb{G}$ by numerical differentiation, as in (3.2). Our more general framework allows us to retain this aspect of the m-out-of-n bootstrap procedure, while using bootstrap samples of size n rather than m to approximate the law of $\mathbb{G}_n$.

4. Uniform confidence bands for F

In this section we consider a variety of methods for constructing uniform confidence bands for F and investigate their performance using numerical simulations. It is apparent from Theorem 2.1 that $(1-\alpha)$-level uniform confidence bands for F may be constructed from $\hat{\mathbb{F}}_n$ and a suitable estimate of the $(1-\alpha)$-quantile of $\|\hat{\mathbb{G}}\|$. Denote this quantile by

$$q^{1-\alpha} = \inf\{c \in \mathbf{R} : P(\|\hat{\mathbb{G}}\| \leq c) \geq 1-\alpha\}.$$

Different estimates of $q^{1-\alpha}$ will yield different confidence bands for F.

The rescaled bootstrap of Dümbgen (1993) discussed in Section 3 provides one way to estimate $q^{1-\alpha}$. The rescaled bootstrap estimate is

$$q_n^{1-\alpha} = \inf\{c \in \mathbf{R} : P(\|\hat{\mathcal{M}}_n' \mathbb{G}_n^*\| \leq c \,|\, \mathcal{X}_n) \geq 1 - \alpha\},$$

which may be computed by evaluating $\|\hat{\mathcal{M}}_n' \mathbb{G}_n^*\|$ at repeated bootstrap realizations of $\mathbb{G}_n^*$. We will call $\hat{\mathbb{F}}_n \pm n^{-1/2} q_n^{1-\alpha}$ the *rescaled bootstrap confidence bands*. With Theorem 3.2 in hand it is straightforward to establish the following result.

Proposition 4.1. *Let Assumptions 2.1, 2.2, 3.1 and 3.2 hold, and suppose that the distribution function of $\|\hat{\mathbb{G}}\|$ is continuous and strictly increasing at $q^{1-\alpha}$. Then $q_n^{1-\alpha} \to q^{1-\alpha}$ in probability, and*

$$P\left(\|\hat{\mathbb{F}}_n - F\| \leq \frac{q_n^{1-\alpha}}{\sqrt{n}}\right) \to 1 - \alpha.$$

A limitation of the rescaled bootstrap (aside from poor small sample performance, which we shall come to shortly) is the need to choose the tuning parameter t_n. There is an alternative procedure we can use to construct conservative confidence bands for F that does not require the use of a tuning parameter. A consequence of the weak contractivity of the LCM operator is that we must have $\|\hat{\mathbb{G}}_n\| \leq \|\mathbb{G}_n\|$ when F is concave. If $\mathbb{F}_n$ is the empirical distribution function of an iid sample drawn from F, and we are willing to assume further that $F(0) = 0$, then the law of $\|\mathbb{G}_n\|$ is distribution free and its $(1-\alpha)$-quantile, which we denote by $\hat{q}_n^{1-\alpha}$, may be obtained by Monte Carlo simulation or from tables of critical values for the Kolmogorov-Smirnov test. This quantile provides an upper bound on the $(1-\alpha)$-quantile of $\|\hat{\mathbb{G}}_n\|$, and so we have

$$P\left(\|\hat{\mathbb{F}}_n - F\| \leq \frac{\hat{q}_n^{1-\alpha}}{\sqrt{n}}\right) \geq 1 - \alpha$$

for any fixed n, without resorting to asymptotics. We will call $\hat{\mathbb{F}}_n \pm n^{-1/2} \hat{q}_n^{1-\alpha}$ the *conservative fixed n confidence bands*.

In more general settings where $\mathbb{F}_n$ is not the empirical distribution function of an iid sample drawn from F, the law of $\|\mathbb{G}_n\|$ may need to be estimated in order to obtain conservative confidence bands for F. Under Assumption 3.1, this may be achieved using a suitable bootstrap technique. Define the quantile

$$\check{q}^{1-\alpha} = \inf\{c \in \mathbf{R} : P(\|\mathbb{G}\| \leq c) \geq 1 - \alpha\}$$

and its bootstrap estimator

$$\check{q}_n^{1-\alpha} = \inf\{c \in \mathbf{R} : P(\|\mathbb{G}_n^*\| \leq c \,|\, \mathcal{X}_n) \geq 1 - \alpha\}.$$

We will call $\hat{\mathbb{F}}_n \pm n^{-1/2} \check{q}_n^{1-\alpha}$ the *conservative bootstrap confidence bands*. The next result confirms they are asymptotically valid though potentially conservative.

Proposition 4.2. *Let Assumptions 2.1, 2.2 and 3.1 hold, and suppose that the distribution function of* $\|\mathbb{G}\|$ *is continuous and strictly increasing at* $q^{1-\alpha}$. *Then* $\check{q}_n^{1-\alpha} \to \check{q}^{1-\alpha}$ *in probability, and*

$$\liminf_{n\to\infty} P\left(\|\hat{\mathbb{F}}_n - F\| \le \frac{\check{q}_n^{1-\alpha}}{\sqrt{n}}\right) \ge 1-\alpha.$$

Our next procedure for constructing confidence bands is motivated by the Delta method approximation (3.1), which suggests the use of bands of the form $\hat{\mathbb{F}}_n \pm n^{-1/2}\tilde{q}_n^{1-\alpha}$ with

$$\tilde{q}_n^{1-\alpha} = \inf\{c \in \mathbf{R} : P(\sqrt{n}\|\mathcal{M}\mathbb{F}_n^* - \hat{\mathbb{F}}_n\| \le c \,|\, \mathcal{X}_n) \ge 1-\alpha\}.$$

We will call $\hat{\mathbb{F}}_n \pm n^{-1/2}\tilde{q}_n^{1-\alpha}$ the *naive bootstrap confidence bands*. As discussed earlier, the Delta method approximation (3.1) is only valid for those F that are strictly concave on their support, because linearity of $\mathcal{M}'_F$ is required to apply the Delta method for the bootstrap. When F is not strictly concave on its support we do not expect the naive bootstrap confidence intervals to achieve their nominal coverage rate asymptotically. Their limiting behavior is instead governed by Theorem 3.1.

The final method of constructing confidence bands we shall consider was suggested by a referee, and has also been studied by Sen et al. (2010). Assume that $\mathbb{F}_n$ is the empirical distribution function of an iid sample $(X_1, \ldots, X_n)$ drawn from F. Let $\mathbb{F}_n^\dagger$ be the empirical distribution function of n iid draws from $\hat{\mathbb{F}}_n$; such draws may be constructed by applying the quantile function corresponding to $\hat{\mathbb{F}}_n$ to each of n draws from the uniform distribution on the unit interval, independent of one another and of $\mathcal{X}_n$, the σ-field generated by $(X_1, \ldots, X_n)$. Define

$$\bar{q}_n^{1-\alpha} = \inf\{c \in \mathbf{R} : P(\sqrt{n}\|\mathcal{M}\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n\| \le c \,|\, \mathcal{X}_n) \ge 1-\alpha\}.$$

We will call $\hat{\mathbb{F}}_n \pm n^{-1/2}\bar{q}_n^{1-\alpha}$ the *constrained bootstrap confidence bands*, because the distribution from which our bootstrap sample is drawn is constrained to satisfy concavity. Their unconditional limiting behavior is governed by the following result.

Theorem 4.1. *Under Assumption 2.1 we have*

$$\sqrt{n}(\mathcal{M}\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n) \rightsquigarrow \mathcal{M}'_F(\mathbb{G} + \mathcal{M}'_F(\mathbb{G}')) - \mathcal{M}'_F(\mathbb{G}'),$$

where $\mathbb{G}'$ *is an independent copy of* $\mathbb{G}$.

Based on the fact that the weak limit in Theorem 4.1 is generally not equal to the weak limit $\hat{\mathbb{G}}$ of $\hat{\mathbb{G}}_n$ unless F is strictly concave, we might expect the constrained confidence bands to perform poorly, but in fact they perform well in our numerical simulations. It turns out that the upper quantiles of the uniform norms of these two weak limits are, as far as we can tell from numerical computations, identical. We will come back to this fortuitous property shortly.

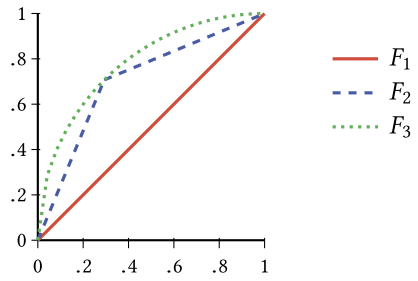


FIG 1. *Distribution functions used in numerical simulations.*

Sen et al. (2010) have shown that confidence intervals for the left-derivative of F at a point obtained by bootstrapping from $\hat{\mathbb{F}}_n$ are asymptotically invalid and have coverage rates well below the nominal level at relevant sample sizes. Moreover, their results indicate that resampling from the Grenander estimator can yield bootstrap processes that weakly converge unconditionally but not conditional on the data; note that Theorem 4.1 says nothing about conditional convergence. The good performance of the constrained confidence bands for F in our numerical simulations is therefore quite surprising.

To assess the small sample behavior of the various confidence bands just described we ran two sets of numerical simulations, with the former involving iid samples and the latter serially dependent samples. First we describe the former set of simulations. In each of 10000 experimental replications we randomly generated an iid sample of size $n = 100$ from one of three distribution functions to be defined shortly, and estimated that distribution using $\hat{\mathbb{F}}_n$, the LCM of the empirical distribution function. Setting $\alpha = 0.1$, we computed the rescaled, naive and constrained bootstrap confidence bands. For the rescaled bootstrap we used twenty values of the tuning parameter t_n running from 0.05 to 1 in increments of 0.05. Note that at $t_n = 0.1 = n^{-1/2}$ the rescaled bootstrap confidence bands are identical to the naive bootstrap confidence bands. All confidence bands were computed using 1000 bootstrap samples. We also computed the conservative fixed n confidence bands, with $\hat{q}_n^{1-\alpha} = 1.2$ obtained by Monte Carlo simulation.

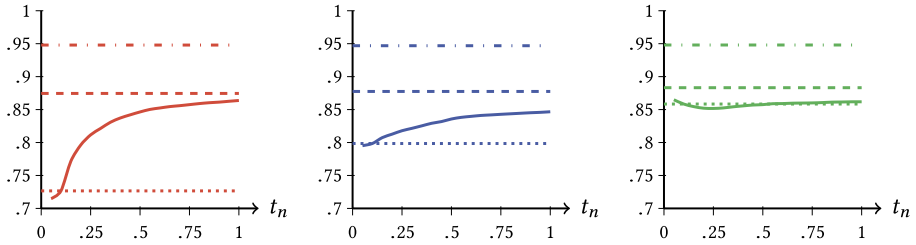
The three distribution functions we used in our numerical simulations are plotted in Figure 1. All three distributions are supported on the unit interval. The first distribution is the uniform distribution on the unit interval, $F_1(x) = x$. The second distribution is the piecewise linear function

$$F_2(x) = \begin{cases} \frac{1}{\sqrt{2}-1}x & \text{if } 0 \le x \le 1 - \frac{1}{\sqrt{2}} \\ 2 - \sqrt{2} + (\sqrt{2}-1)x & \text{if } 1 - \frac{1}{\sqrt{2}} \le x \le 1. \end{cases}$$

The third distribution, whose graph forms a quarter circle, is

$$F_3(x) = \sqrt{x(2-x)}.$$

Note that F_1 is linear on its support, that F_3 is strictly concave on its support, and that F_2 is concave but not strictly concave on its support.



(a) F_1 (b) F_2 (c) F_3

FIG 2. *Coverage rates under iid sampling. Solid, dashed and dotted lines show coverage rates for the rescaled, constrained and naive bootstrap confidence bands respectively. Dash-dotted lines show coverage rates for the conservative fixed n confidence bands.*

The results of our first set of numerical simulations are displayed in Figure 2. The three panels correspond to the three different distribution functions plotted in Figure 1. The vertical axes measure coverage rates and the horizontal axes indicate the tuning parameter t_n used for the rescaled bootstrap. Coverage rates for the rescaled, naive and constrained bootstrap confidence bands are plotted with solid, dotted and dashed lines respectively. Coverage rates for the conservative fixed n confidence bands are plotted with dash-dotted lines. We enumerate several observations on the displayed results.

1. Theorem 3.1 implies that the naive bootstrap confidence bands should have limiting coverage rate equal to the nominal level for the strictly concave distribution function F_3, and indeed the computed coverage rate with $n = 100$ is 0.86, fairly close to the nominal level of 0.9. The coverage rates are 0.73 and 0.80 for F_1 and F_2 respectively, well below the nominal level.
2. The conservative fixed n confidence bands are, as expected, conservative, with coverage rates of 0.95 for F_1, F_2 and F_3.
3. The rescaled bootstrap confidence bands do not perform well, despite the asymptotic justification for their use provided by Proposition 4.1. At $t_n = n^{-1/2} = 0.1$ they are identical to the naive bootstrap confidence bands by construction. For F_1 and F_2, as we increase t_n the coverage rate rises, but never attains the nominal rate. For F_3 the coverage rate is insensitive to the choice of t_n. Additional unreported simulations showed the coverage rate for all three distributions effectively flat for $t_n > 1$.
4. The constrained bootstrap confidence bands perform well, with coverage rates of 0.87, 0.88 and 0.88 for F_1, F_2 and F_3 respectively. They dominate the rescaled bootstrap confidence bands across all values of t_n.

The good performance of the constrained bootstrap confidence bands is surprising because Theorem 4.1 indicates that the weak limit of $\sqrt{n}(\mathcal{M}\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n)$ differs from $\hat{\mathbb{G}}$ except when F is strictly concave. To investigate further, we computed the quantile functions of the uniform norms of the weak limits $\hat{\mathbb{G}}$, $\mathcal{M}'_F(\mathbb{G} + \mathbb{G}') - \mathcal{M}'_F(\mathbb{G}')$ and $\mathcal{M}'_F(\mathbb{G} + \mathcal{M}'_F(\mathbb{G}')) - \mathcal{M}'_F(\mathbb{G}')$ appearing in The-

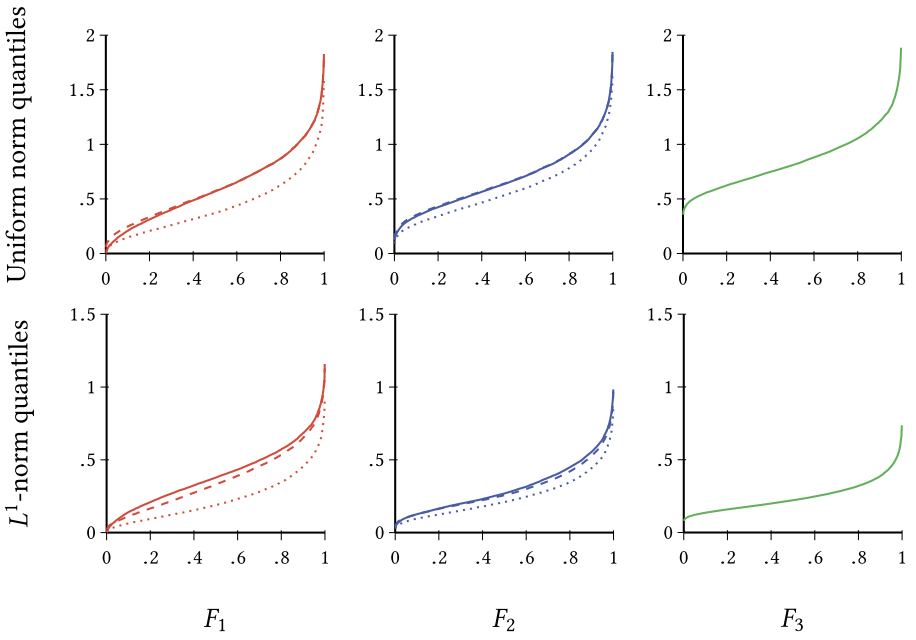


FIG 3. *Quantile functions for the uniform and L^1-norms of the weak limits $\hat{\mathbb{G}}$, $\mathcal{M}'_F(\mathbb{G}+\mathbb{G}')-\mathcal{M}'_F(\mathbb{G}')$ and $\mathcal{M}'_F(\mathbb{G}+\mathcal{M}'_F(\mathbb{G}'))-\mathcal{M}'_F(\mathbb{G}')$ are shown in solid, dotted and dashed lines respectively.*

orems 2.1, 3.1 and 4.1, for each of the three distributions F_1, F_2 and F_3. This was achieved by ordering the uniform norms of 10000 independent realizations of the three weak limits; each such realization was constructed from two independent Brownian bridges $\mathbb{B}$ and $\mathbb{B}'$ computed using Karhunen-Loève series truncated at 10000 terms, evaluated over a grid of 1000 points spread evenly over the unit interval. The computed quantile functions are displayed in the top three panels of Figure 3. The three panels correspond to F_1, F_2 and F_3 respectively. Solid, dotted and dashed lines are used to plot the quantile functions of the uniform norms of the weak limits $\hat{\mathbb{G}}$, $\mathcal{M}'_F(\mathbb{G}+\mathbb{G}')-\mathcal{M}'_F(\mathbb{G}')$ and $\mathcal{M}'_F(\mathbb{G}+\mathcal{M}'_F(\mathbb{G}'))-\mathcal{M}'_F(\mathbb{G}')$ respectively. As expected, the three quantile functions coincide in the third panel, because the weak limits coincide when F is strictly concave. The quantile function for $\|\mathcal{M}'_F(\mathbb{G}+\mathbb{G}')-\mathcal{M}'_F(\mathbb{G}')\|$ lies below the quantile function for $\|\hat{\mathbb{G}}\|$ in the first two panels, consistent with the undercoverage of the naive bootstrap confidence bands shown in the first two panels of Figure 2. What is more interesting is that the quantile functions for $\|\hat{\mathbb{G}}\|$ and $\|\mathcal{M}'_F(\mathbb{G}+\mathcal{M}'_F(\mathbb{G}'))-\mathcal{M}'_F(\mathbb{G}')\|$ are nearly identical in the first two panels. Visible discrepancies occur only at quantiles of around 0.4 and below. The near-equality of these two quantile functions explains the good behavior of the constrained bootstrap confidence bands in Figure 2. Note that while the laws of the uniform norms of $\hat{\mathbb{G}}$ and $\mathcal{M}'_F(\mathbb{G}+\mathcal{M}'_F(\mathbb{G}'))-\mathcal{M}'_F(\mathbb{G}')$ are approxi-

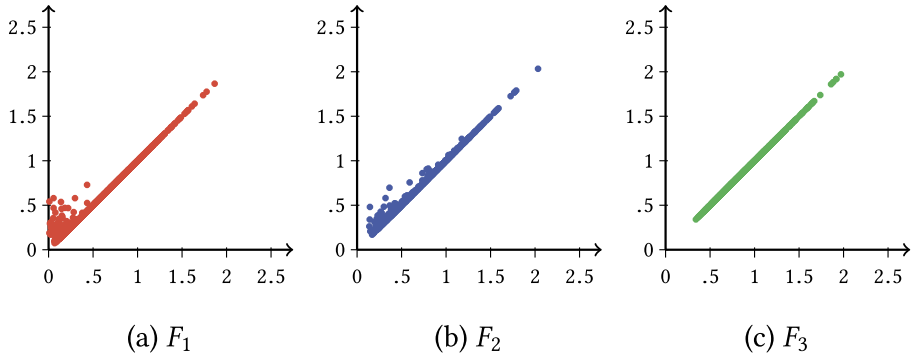


FIG 4. *Scatterplots of* $\|\hat{\mathbb{G}}\|$ *versus* $\|\mathcal{M}'_F(\mathbb{G} + \mathcal{M}'_F(\mathbb{G}')) - \mathcal{M}'_F(\mathbb{G}')\|$.

mately equal, the same is not true more generally for other choices of norm. In the second row of panels in Figure 3 we plot quantile functions for the L^1-norms of the three weak limits. The three quantile functions are more clearly distinct from one another in the first two of these panels.

To provide further insight into the relationship between $\|\hat{\mathbb{G}}\|$ and $\|\mathcal{M}'_F(\mathbb{G} + \mathcal{M}'_F(\mathbb{G}')) - \mathcal{M}'_F(\mathbb{G}')\|$, in Figure 4 we display scatterplots of the two quantities based on 1000 independent draws of $(\mathbb{G}, \mathbb{G}')$. The horizontal axes measure $\|\hat{\mathbb{G}}\|$ and the vertical axes measure $\|\mathcal{M}'_F(\mathbb{G} + \mathcal{M}'_F(\mathbb{G}')) - \mathcal{M}'_F(\mathbb{G}')\|$. In panel (c), all realizations lie on the 45°-line, because $\mathcal{M}'_F$ is linear (and idempotent). In panels (a) and (b) this is not the case; however, deviations from the 45° line occur overwhelmingly toward the lower quantiles of the distributions of the two uniform norms. In panel (a) we do not see a single deviation from the 45° line beyond 0.5 on the horizontal axis, which corresponds roughly to the 0.4 quantile of $\|\hat{\mathbb{G}}\|$.

The poor performance of the rescaled bootstrap confidence bands apparent in Figure 2 was anticipated by Dümbgen (1993, p. 126), who warned that the rescaled bootstrap is "very nonrobust and shouldn't be used". He attributed this nonrobustness to the fact that the rescaled bootstrap typically fails to be consistent under a sequence of parameters indexed by the sample size (in our case, under a sequence of distributions F_n) converging at a rate no faster than $n^{-1/2}$. Note that the consistency property we established in Theorem 3.2 requires F to be fixed. The results in Figure 2 confirm the pertinence of Dümbgen's warning, especially in view of the dominant performance of the constrained bootstrap confidence bands. In additional unreported simulations, we investigated whether the performance of the rescaled bootstrap improved with larger sample sizes. We found that its performance continued to be disappointing even with sample sizes as large as $n = 1000$, exhibiting substantial undercoverage over all tuning parameter values. By comparison, the constrained bootstrap confidence bands obtained a coverage rate of nearly exactly 0.9 with $n = 1000$.

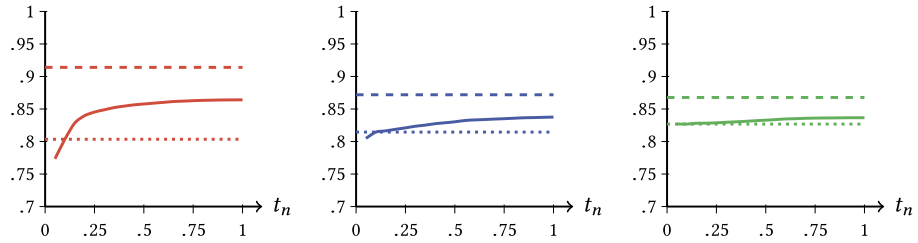


FIG 5. *Coverage rates under dependent sampling. Solid, dashed and dotted lines show coverage rates for the rescaled, conservative and naive bootstrap confidence bands respectively.*

We ran a second set of simulations to investigate the case of dependent sampling. The samples generated were stationary Markov chains of length $n = 100$. The invariant distribution of each chain was chosen to be F_1, F_2 or F_3, as displayed in Figure 1. Dependence in each chain was induced using a Gaussian copula function with correlation parameter $\rho = 0.75$. This means that the bivariate distribution of any two consecutive observations in a chain has Gaussian copula with $\rho = 0.75$ and marginal distributions both equal to one of F_1, F_2 and F_3. See Beare (2010, 2012) for more on copula-based Markov models and their dependence properties. For each Markov chain of length n we computed its empirical distribution function $\mathbb{F}_n$. Bootstrap versions $\mathbb{F}_n^*$ of $\mathbb{F}_n$ were generated by applying the so-called local bootstrap of Paparoditis and Politis (2002), which resamples observations in such a way as to preserve Markovian dependence. A tuning parameter used for the local bootstrap was chosen using a plug-in procedure based on an auxiliary first-order autoregression, as described by Paparoditis and Politis (2002, p. 315). See Beare and Seo (2014) for additional discussion of the local bootstrap and its application to copula-based Markov models.

The results of our second set of simulations are displayed in Figure 5. As with the first set of simulations, coverage rates were computed over 10000 experimental replications, and confidence bands were computed using 1000 bootstrap samples. Coverage rates were computed for the rescaled, conservative and naive bootstrap confidence bands, all based on the local bootstrap. We did not compute coverage rates for the constrained bootstrap confidence bands or conservative fixed n confidence bands because their applicability is confined to the iid setting. We enumerate some brief comments on the displayed results.

1. The naive bootstrap confidence bands have coverage rates of 0.80 and 0.81 for F_1 and F_2 respectively, which is an improvement on the iid case, but nevertheless well below the nominal level of 0.9. For the strictly concave distribution F_3, where Theorem 3.1 indicates that we might expect the naive confidence bands to perform well, the coverage rate is only 0.83.
2. Despite the asymptotic conservatism established in Proposition 4.2, the conservative bootstrap confidence bands are in fact only conservative for

F_1. For F_2 and F_3 their coverage rate is 0.87, a little below the nominal level.

3. The coverage rates for the rescaled bootstrap confidence bands are qualitatively similar to the iid case, being equal to the coverage rate of the naive bootstrap confidence bands at $t_n = n^{-1/2} = 0.1$ and then either rising (for F_1 and F_2) or staying relatively flat (for F_3) as t_n increases. They are dominated by the conservative bootstrap confidence bands for all values of t_n.

We conclude from the results displayed in Figure 5 that the conservative bootstrap confidence bands are the most reliable of the three confidence bands considered. These results are of course specific to the structure of dependence and method of resampling considered, but they at least point to the possibility of effective inference in a dependent setting using the conservative bootstrap confidence bands, and provide further evidence of poor performance of the rescaled and naive bootstrap confidence bands.

5. Proofs

Proof of Proposition 2.1. Weak contractivity of $\mathcal{M}$ yields $t_n^{-1}\|\mathcal{M}(\theta + t_n h_n) - \mathcal{M}(\theta + t_n h)\| \le \|h_n - h\| \to 0$. It therefore suffices for us to demonstrate Gâteaux, rather than Hadamard, directional differentiability of $\mathcal{M}$. That is, we wish to show that for each concave $\theta \in \ell^\infty$, the maps $\mathcal{M}'_{\theta,n} : C_0 \to \ell^\infty$ and $\zeta_\theta : C_0 \to \ell^\infty$ defined by

$$\mathcal{M}'_{\theta,n} h = t_n^{-1}[\mathcal{M}(\theta + t_n h) - \mathcal{M}(\theta)] \quad \text{for each } n \in \mathbf{N} \text{ and each } h \in C_0$$

and

$$\zeta_\theta h(x) = \mathcal{M}_{T_{\theta,x}} h(x) \quad \text{for each } h \in C_0 \text{ and each } x \in \mathbf{R}_+$$

satisfy

$$\sup_{x \in \mathbf{R}_+} |\mathcal{M}'_{\theta,n} h(x) - \zeta_\theta h(x)| \to 0 \quad \text{for each } h \in C_0. \tag{5.1}$$

Under (5.1) we may conclude that $\mathcal{M}$ is Hadamard directionally differentiable at θ tangentially to C_0, with directional derivative $\mathcal{M}'_\theta = \zeta_\theta$. We shall verify (5.1) for all concave $\theta \in \ell^\infty$ by applying Dini's theorem. This involves four steps. First we show that $\mathcal{M}'_{\theta,n} h$ and $\zeta_\theta h$ are continuous on $\mathbf{R}^+$. Next we find continuous extensions of $\mathcal{M}'_{\theta,n} h$ and $\zeta_\theta h$ to $\bar{\mathbf{R}}^+ := [0, \infty]$, the one point compactification of $\mathbf{R}^+$. Next we show that $\mathcal{M}'_{\theta,n} h(x)$ is nonincreasing in n for each $x \in \bar{\mathbf{R}}^+$. Finally we show that $\mathcal{M}'_{\theta,n} h(x) \to \zeta_\theta h(x)$ for each $x \in \bar{\mathbf{R}}^+$.

To show that $\mathcal{M}'_{\theta,n} h$ is continuous on $\mathbf{R}^+$, we first observe that continuity of $\mathcal{M}'_{\theta,n} h$ on $(0, \infty)$ follows from the fact that the LCMs $\mathcal{M}(\theta)$ and $\mathcal{M}(\theta + t_n h)$ are both concave on $\mathbf{R}^+$, hence continuous on $(0, \infty)$. It remains to establish continuity at zero. Since $\mathcal{M}(\theta)$ is bounded from below and nondecreasing on $(0, \infty)$ (it is bounded from below by $-\|\theta\|$ and must therefore be nondecreasing due to its concavity), its right-limit at zero exists, and we denote it by $\mathcal{M}(\theta)(0+)$.

The right-limit of θ at zero exists for the same reason, and we denote it by $\theta(0+)$. We will show that

$$\mathcal{M}(\theta)(0+) = \theta(0+) \tag{5.2}$$

by supposing that this equality is false and deducing a contradiction. Due to the fact that $\mathcal{M}(\theta)$ majorizes θ, if (5.2) is false then there exists $\epsilon > 0$ and $a < \mathcal{M}(\theta)(0+)$ such that, for all $x \in (0, \epsilon)$, we have $\mathcal{M}(\theta)(x) > a > \theta(x)$. Let $\vartheta \in \ell^\infty$ be equal to $\mathcal{M}(\theta)$ on $[\epsilon, \infty)$, be given by

$$\vartheta(x) = a + \epsilon^{-1}(\mathcal{M}(\theta)(\epsilon) - a)x$$

for $x \in (0, \epsilon)$, and be equal to $\theta(0)$ at zero. Then ϑ is a concave majorant of θ lying strictly below $\mathcal{M}(\theta)$ on $(0, \epsilon)$, a contradiction. This establishes (5.2), and a similar argument using $\theta + t_n h$ in place of θ establishes that

$$\mathcal{M}(\theta + t_n h)(0+) = \theta(0+) + t_n h(0). \tag{5.3}$$

At zero the LCMs satisfy $\mathcal{M}(\theta)(0) = \theta(0)$ and $\mathcal{M}(\theta + t_n h)(0) = \theta(0) + t_n h(0)$. Combining these equalities with (5.2) and (5.3), we obtain

$$\begin{aligned}\lim_{x \downarrow 0} \mathcal{M}'_{\theta,n} h(x) &= t_n^{-1}[\mathcal{M}(\theta + t_n h)(0+) - \mathcal{M}(\theta)(0+)] \\ &= t_n^{-1}[\mathcal{M}(\theta + t_n h)(0) - \mathcal{M}(\theta)(0)] \\ &= \mathcal{M}'_{\theta,n} h(0).\end{aligned}$$

This shows that $\mathcal{M}'_{\theta,n} h$ is continuous at zero, and hence everywhere on $\mathbf{R}^+$.

To show that $\zeta_\theta h$ is continuous on $\mathbf{R}^+$ observe that for any $x \in \mathbf{R}^+$ either the set $T_{\theta,x}$ is an open interval containing x, or it is the singleton $\{x\}$. In the former case $\zeta_\theta h$ coincides with the concave function $\mathcal{M}_{T_{\theta,x}} h$ over $T_{\theta,x}$, so $\zeta_\theta h$ is continuous at x. Suppose instead that $T_{\theta,x} = \{x\}$, so that $\zeta_\theta h(x) = h(x)$ (this is the relevant case when $x = 0$). Then for any $x' > x$ it must be the case that $T_{\theta,x'} \subseteq (x, \infty)$, and consequently

$$h(x') \leq \zeta_\theta h(x') \leq \mathcal{M}_{(x,\infty)} h(x'). \tag{5.4}$$

The function h is continuous and therefore has right-limit $h(x)$ at x. We can show that the function $\mathcal{M}_{(x,\infty)} h$ has right-limit $h(x)$ at x in the same way that we showed (5.2). Letting x' decrease to x, we deduce from (5.4) that $\zeta_\theta h$ has right-limit $h(x)$ at x, and is therefore right-continuous at x. A symmetric argument shows that $\zeta_\theta h$ is left-continuous at x when $x > 0$. We conclude that $\zeta_\theta h$ is continuous everywhere on $\mathbf{R}^+$, as claimed.

Next we obtain continuous extensions of $\mathcal{M}_{\theta,n} h$ and $\zeta_\theta h$ to $\bar{\mathbf{R}}^+$. A bounded, nondecreasing real valued function on $\mathbf{R}^+$—in particular, any concave member of ℓ^∞—is convergent to its supremum at infinity. Since $\mathcal{M}'_{\theta,n} h$ is t_n^{-1} times the difference of two such functions, $\mathcal{M}(\theta + t_n h)$ and θ, we may extend it continuously to $\bar{\mathbf{R}}^+$ by defining

$$\begin{aligned}\mathcal{M}'_{\theta,n} h(\infty) &= t_n^{-1}\left[\lim_{x \to \infty} \mathcal{M}(\theta + t_n h)(x) - \lim_{x \to \infty} \theta(x)\right] \\ &= t_n^{-1}\left[\sup_{x \in \mathbf{R}^+} (\theta(x) + t_n h(x)) - \sup_{x \in \mathbf{R}^+} \theta(x)\right].\end{aligned} \tag{5.5}$$

To obtain a continuous extension of $\zeta_\theta h$ to $\bar{\mathbf{R}}^+$, first consider the case where $a := \inf\{x \in \mathbf{R}^+ : \theta(x) = \sup_{y\in\mathbf{R}^+} \theta(y)\} < \infty$, so that θ achieves its supremum on $[a, \infty)$, or possibly $(0, \infty)$ if $a = 0$. In this case we have $\zeta_\theta h(x) = \mathcal{M}_{(a,\infty)}h(x)$ for all $x > a$. The function $\mathcal{M}_{(a,\infty)}h$ is concave and bounded from below by h, which vanishes at infinity, so $\mathcal{M}_{(a,\infty)}h$ must be nondecreasing. Also, $\mathcal{M}_{(a,\infty)}h$ is bounded from above by $\|h\|$. Since $\mathcal{M}_{(a,\infty)}h$ is nondecreasing and bounded it must be convergent at infinity, and so we may continuously extend $\zeta_\theta h$ to $\bar{\mathbf{R}}^+$ by defining

$$\zeta_\theta h(\infty) = \lim_{x\to\infty} \mathcal{M}_{(a,\infty)}h(x) = \sup_{x\geq a} h(x).$$

Next suppose that we instead have $a = \infty$. In this case θ does not achieve its supremum on $\mathbf{R}^+$, and consequently $T_{\theta,x}$ is of finite length for each $x \in \mathbf{R}^+$. Fix $\epsilon > 0$. Since $h \in C_0$, there exists $N < \infty$ such that, for all $x \geq N$, we have $|h(x)| < \epsilon$. But $T_{\theta,N}$ is of finite length, so there must exist a finite $N' > N$ such that N is less than all elements of $T_{\theta,N'}$. For all $x \geq N'$ we have $|\zeta_\theta h(x)| = |\mathcal{M}_{T_{\theta,x}}h(x)| \leq |\mathcal{M}_{[N,\infty)}h(x)| < \epsilon$. Since ϵ was arbitrary it follows that $\zeta_\theta h$ vanishes at infinity. We thus obtain a continuous extension of $\zeta_\theta h$ to $\bar{\mathbf{R}}^+$ in the case $a = \infty$ by defining $\zeta_\theta h(\infty) = 0$.

Before continuing further it will be useful to extend the domain and codomain of the LCM operator to accommodate sums of bounded functions and affine functions. Letting S denote the collection of maps from $\mathbf{R}^+$ to $\mathbf{R}$ formed by summing a function in ℓ^∞ and an affine function, we define

$$\mathcal{M}f(x) = \inf\{g(x) : g \in S,\ g \text{ is concave, and } f \leq g\}, \quad f \in S, \quad x \in \mathbf{R}^+,$$

which evidently reduces to our earlier definition of $\mathcal{M}f$ when $f \in \ell^\infty$. Consider functions $f, g \in S$ with g affine, and write the latter as $g(x) = \sigma + \tau x$ with $\sigma, \tau \in \mathbf{R}$. The second part of the first sentence of Lemma 2.1 of Durot and Tocquet (2003) implies that $\mathcal{M}(f + g)(x) = \mathcal{M}(f + \sigma)(x) + \tau x$, and the first part of the same sentence implies that $\mathcal{M}(f+\sigma) = \mathcal{M}f+\sigma$. This proves a useful property of $\mathcal{M}$: we have $\mathcal{M}(f + g) = \mathcal{M}f + g$ for any $f, g \in S$ with g affine.

We next show that $\mathcal{M}'_{\theta,n}h(x)$ is nonincreasing in n for each $x \in \bar{\mathbf{R}}^+$. First consider the case $x \in (0, \infty)$; our arguments for this case are similar to those in the proof of Lemma 3.2 of Beare and Moon (2015). Since θ is concave, for each $x \in (0, \infty)$ the supporting hyperplane theorem ensures the existence of an affine function $\xi_x : \mathbf{R}^+ \to \mathbf{R}$ such that $\xi_x(x) = \theta(x)$ and $\xi_x \geq \theta$. We now have

$$\begin{aligned}
\mathcal{M}'_{\theta,n}h(x) &= t_n^{-1}[\mathcal{M}(\theta + t_n h)(x) - \xi_x(x)] && \text{since } \theta \text{ is concave and } \xi_x(x) = \theta(x) \\
&= t_n^{-1}\mathcal{M}(\theta + t_n h - \xi_x)(x) && \text{since } -\xi_x \text{ is affine} \\
&= \mathcal{M}(h + t_n^{-1}(\theta - \xi_x))(x) && \text{since } \mathcal{M} \text{ is positive homogeneous.}
\end{aligned}$$

Since $\mathcal{M}$ is monotone and $\xi_x \geq \theta$, this shows that $\mathcal{M}'_{\theta,n}h(x)$ is nonincreasing in n for each $x \in (0, \infty)$. Next consider the case $x = 0$. Since $\mathcal{M}f(0) = f(0)$ for every $f \in S$, we have $\mathcal{M}'_{\theta,n}h(0) = h(0)$, which is constant and hence nondecreasing in n. Finally consider the case $x = \infty$. Letting $\xi_\infty = \sup_{x\in\mathbf{R}^+} \theta(x)$, we see from (5.5) that $\mathcal{M}'_{\theta,n}h(\infty) = \sup_{x\in\mathbf{R}^+}[h(x)+t_n^{-1}(\theta(x)-\xi_\infty)]$, which is nondecreasing in n.

It remains to demonstrate that $\mathcal{M}'_{\theta,n}h(x) \to \zeta_\theta h(x)$ for each $x \in \bar{\mathbf{R}}^+$. It is clear that $\zeta_\theta h(0) = h(0)$ and we observed already that $\mathcal{M}'_{\theta,n}h(0) = h(0)$ for every n, so the desired convergence holds for $x = 0$. From (5.5), we also have

$$\mathcal{M}'_{\theta,n}h(\infty) = t_n^{-1}\left[\sup_{x\in\mathbf{R}^+}(\theta(x) + t_n h(x)) - \sup_{x\in\mathbf{R}^+}\theta(x)\right] \to \sup_{x\in\arg\max\theta} h(x),$$

with the stated convergence following from the Hadamard directional differentiability of the supremum operator; see e.g. Lemma B.1 of Fang and Santos (2015), which applies since $\bar{\mathbf{R}}^+$ is homeomorphic to $[0,1]$. Note that we should interpret $h(\infty) = 0$ and $\arg\max\theta = \{\infty\}$ if θ does not attain its supremum on $\mathbf{R}^+$. We now have $\mathcal{M}'_{\theta,n}h(\infty) \to \sup_{x\geq a} h(x)$ if θ attains its supremum on $\mathbf{R}^+$, and $\mathcal{M}'_{\theta,n}h(\infty) \to 0$ otherwise, so that in either case we have the desired convergence $\mathcal{M}'_{\theta,n}h(\infty) \to \zeta_\theta h(\infty)$. It remains to establish convergence at points $x \in (0,\infty)$. Our arguments here are again similar to those in the proof of Lemma 3.2 of Beare and Moon (2015). Let $h_{\theta,n,x} = h + t_n^{-1}(\theta - \xi_x)$, and recall that $\mathcal{M}'_{\theta,n}h(x) = \mathcal{M}h_{\theta,n,x}(x)$. A representation of the LCM in terms of a supremum of secant segments (cf. Carolan, 2002, Lem. 1) allows us to write

$$\mathcal{M}h_{\theta,n,x}(x) = \sup_{x'\in[0,x)}\sup_{x''\in(x,\infty)}\left[\frac{x''-x}{x''-x'}h_{\theta,n,x}(x') + \frac{x-x'}{x''-x'}h_{\theta,n,x}(x'')\right]. \quad (5.6)$$

Since ξ_x is affine and $\xi_x(x) = \theta(x)$, the term in square brackets in (5.6) equals

$$\frac{x''-x}{x''-x'}h(x') + \frac{x-x'}{x''-x'}h(x'') + t_n^{-1}\left[\frac{x''-x}{x''-x'}\theta(x') + \frac{x-x'}{x''-x'}\theta(x'') - \theta(x)\right]. \quad (5.7)$$

The term in square brackets in (5.7) is equal to zero whenever $x', x'' \in T_{\theta,x}$ and is less than zero otherwise. Moreover, letting $T^\delta_{\theta,x} = \{y \in \mathbf{R}^+ : |y - z| < \delta \text{ for some } z \in T_{\theta,x}\}$, for any $\delta > 0$ there exists $\epsilon > 0$ such that the term in square brackets in (5.7) is less than $-\epsilon$ whenever we do not have $x', x'' \in T^\delta_{\theta,x}$. For n large enough that $\sup_{y\in\mathbf{R}^+} h(y) - t_n^{-1}\epsilon < \inf_{y\in\mathbf{R}^+} h(y)$, we may therefore exclude points outside of $T^\delta_{\theta,x}$ from the suprema in (5.6), yielding

$$\begin{aligned}
&\mathcal{M}h_{\theta,n,x}(x) \\
&\quad= \sup_{x'\in[0,x)\cap T^\delta_{\theta,x}}\sup_{x''\in(x,\infty)\cap T^\delta_{\theta,x}}\left[\frac{x''-x}{x''-x'}h_{\theta,n,x}(x') + \frac{x-x'}{x''-x'}h_{\theta,n,x}(x'')\right] \\
&\quad\leq \sup_{x'\in[0,x)\cap T^\delta_{\theta,x}}\sup_{x''\in(x,\infty)\cap T^\delta_{\theta,x}}\left[\frac{x''-x}{x''-x'}h(x') + \frac{x-x'}{x''-x'}h(x'')\right] = \mathcal{M}_{T^\delta_{\theta,x}}h(x),
\end{aligned}$$

with the inequality following from the nonpositivity of the term in square brackets in (5.7). Since $h_{\theta,n,x} = h$ on $T_{\theta,x}$, we also have $\mathcal{M}_{T_{\theta,x}}h(x) \leq \mathcal{M}h_{\theta,n,x}(x)$. We have now shown that $\mathcal{M}_{T_{\theta,x}}h(x) \leq \mathcal{M}'_{\theta,n}h(x) \leq \mathcal{M}_{T^\delta_{\theta,x}}h(x)$ for all n sufficiently large, implying that

$$\limsup_{n\to\infty}|\mathcal{M}'_{\theta,n}h(x) - \mathcal{M}_{T_{\theta,x}}h(x)| \leq |\mathcal{M}_{T^\delta_{\theta,x}}h(x) - \mathcal{M}_{T_{\theta,x}}h(x)|.$$

To obtain the desired convergence $\mathcal{M}'_{\theta,n}h(x) \to \mathcal{M}_{T_{\theta,x}}h(x) = \zeta_\theta h(x)$ it remains only to show that $\mathcal{M}_{T^\delta_{\theta,x}}h(x) \to \mathcal{M}_{T_{\theta,x}}h(x)$ as $\delta \to 0$. Fix $\eta > 0$. Since $h \in C_0$ it is uniformly continuous on $\mathbf{R}^+$, and so for all sufficiently small $\delta > 0$ we have $|h(x') - h(x'')| < \eta$ whenever $|x' - x''| < \delta$. Therefore, using the representation of an LCM as a supremum of secant segments once more, we have

$$\begin{aligned}
\mathcal{M}_{T^\delta_{\theta,x}}h(x) &= \sup_{x' \in [0,x) \cap T^\delta_{\theta,x}} \sup_{x'' \in (x,\infty) \cap T^\delta_{\theta,x}} \left[\frac{x'' - x}{x'' - x'} h(x') + \frac{x - x'}{x'' - x'} h(x'') \right] \\
&\le \sup_{x' \in [0,x) \cap T_{\theta,x}} \sup_{x'' \in (x,\infty) \cap T_{\theta,x}} \left[\frac{x'' - x}{x'' - x'} h(x') + \frac{x - x'}{x'' - x'} h(x'') \right] + \eta \\
&= \mathcal{M}_{T_{\theta,x}}h(x) + \eta
\end{aligned}$$

for all sufficiently small $\delta > 0$. Since $\eta > 0$ was arbitrary and $\mathcal{M}_{T_{\theta,x}}h(x) \le \mathcal{M}_{T^\delta_{\theta,x}}h(x)$, we conclude that $\mathcal{M}_{T^\delta_{\theta,x}}h(x) \to \mathcal{M}_{T_{\theta,x}}h(x)$ as $\delta \to 0$. □

Proof of Proposition 2.2. Weak contractivity of $\mathcal{M}$ implies that

$$\|\mathcal{M}'_{\theta,n}h_1 - \mathcal{M}'_{\theta,n}h_2\| = t_n^{-1}\|\mathcal{M}(\theta + t_n h_1) - \mathcal{M}(\theta + t_n h_2)\| \le \|h_1 - h_2\|.$$

Letting $n \to \infty$ proves weak contractivity of $\mathcal{M}'_\theta$. Next, we use the convexity of $\mathcal{M}$ to write

$$\begin{aligned}
\mathcal{M}'_{\theta,n}(\alpha h_1 + (1-\alpha)h_2) &= t_n^{-1}[\mathcal{M}(\alpha(\theta + t_n h_1) + (1-\alpha)(\theta + t_n h_2)) - \mathcal{M}(\theta)] \\
&\le t_n^{-1}[\alpha\mathcal{M}(\theta + t_n h_1) + (1-\alpha)\mathcal{M}(\theta + t_n h_2) - \mathcal{M}(\theta)] \\
&= \alpha\mathcal{M}'_{\theta,n}h_1 + (1-\alpha)\mathcal{M}'_{\theta,n}h_2.
\end{aligned}$$

Letting $n \to \infty$ proves convexity of $\mathcal{M}'_\theta$. Finally, we observe that if θ is strictly concave on $\mathbf{R}^+$ then $T_{\theta,x}$ is a singleton for all $x \in \mathbf{R}^+$, and so $\mathcal{M}'_\theta$ is the identity. If instead θ is affine in a neighborhood of some $x \in \mathbf{R}^+$ then $\mathcal{M}_{T_{\theta,x}}h \ne -\mathcal{M}_{T_{\theta,x}}(-h)$ unless h is constant on $T_{\theta,x}$, and hence $\mathcal{M}'_\theta$ is not linear. □

Proof of Theorem 2.1. In view of Proposition 2.1, this is immediate from the Delta method for Hadamard directionally differentiable operators, as stated by Shapiro (1991, Thm. 2.1), Dümbgen (1993, Prop. 1) or more recently Fang and Santos (2015, Thm. 2.1). □

Proof of Theorem 3.1. In view of Proposition 2.1, this is immediate from Proposition 1 of Dümbgen (1993). Alternatively, we could appeal to Theorem 2.2 of Kosorok (2008a) to obtain

$$(\sqrt{n}(\mathbb{F}_n^* - \mathbb{F}_n), \sqrt{n}(\mathbb{F}_n - F)) \rightsquigarrow (\mathbb{G}, \mathbb{G}') \quad \text{in } \ell^\infty \times \ell^\infty,$$

and then to the continuous mapping theorem to obtain

$$\sqrt{n}((\mathbb{F}_n^*, \mathbb{F}_n) - (F, F)) \rightsquigarrow (\mathbb{G} + \mathbb{G}', \mathbb{G}') \quad \text{in } \ell^\infty \times \ell^\infty,$$

and then finally apply the functional Delta method with the map

$$\ell^\infty \times \ell^\infty \ni (\theta_1, \theta_2) \mapsto \mathcal{M}(\theta_1) - \mathcal{M}(\theta_2) \in \ell^\infty$$

to obtain the desired weak convergence in ℓ^∞. □

Proof of Theorem 3.2. In view of Proposition 2.1, this is immediate from Proposition 2 of Dümbgen (1993). In the interests of clarity it might also be helpful to prove our result by applying Theorem 3.2 of Fang and Santos (2015), the assumptions of which we now verify. Assumptions 2.1(i), 3.1(i) and 3.2 of Fang and Santos are satisfied by construction. Assumption 2.1(ii) of Fang and Santos is implied by Proposition 2.1. Assumption 2.2 of Fang and Santos is implied by our Assumptions 2.1 and 2.2. Assumption 2.1(iii) of Fang and Santos is satisfied since C is closed (see Remark 2.2 of Fang and Santos). Assumption 3.1(ii) of Fang and Santos is implied by our Assumption 3.1. It remains to verify Assumption 3.3 of Fang and Santos. In view of Remark 3.5 of Fang and Santos, it suffices for us to show that (1) $\hat{\mathcal{M}}'_n$ is Lipschitz continuous uniformly in n, and (2) $\|\hat{\mathcal{M}}'_n h - \mathcal{M}'_F h\| \to 0$ in outer probability for each $h \in C$. (1) is true since

$$\|\hat{\mathcal{M}}'_n h_1 - \hat{\mathcal{M}}'_n h_2\| = t_n^{-1}\|\mathcal{M}(\mathbb{F}_n + t_n h_1) - \mathcal{M}(\mathbb{F}_n + t_n h_2)\| \leq \|h_1 - h_2\|,$$

by weak contractivity of $\mathcal{M}$. To prove (2) we write

$$\hat{\mathcal{M}}'_n h = t_n^{-1}[\mathcal{M}(F + t_n h_n) - \mathcal{M}F] - (\sqrt{n}t_n)^{-1} \cdot \sqrt{n}(\mathcal{M}\mathbb{F}_n - \mathcal{M}F),$$

where $h_n = h + (\sqrt{n}t_n)^{-1}\sqrt{n}(\mathbb{F}_n - F)$. Since $\|h_n - h\| \to 0$ in outer probability under Assumptions 2.2 and 3.2, we have $\|t_n^{-1}[\mathcal{M}(F+t_n h_n) - \mathcal{M}F] - \mathcal{M}'_F h\| \to 0$ in outer probability by Proposition 2.1 and the extended continuous mapping theorem (see e.g. Dümbgen, 1993, p. 136). We also have $\|(\sqrt{n}t_n)^{-1} \cdot \sqrt{n}(\mathcal{M}\mathbb{F}_n - \mathcal{M}F)\| \to 0$ in outer probability by Assumption 3.2 and Theorem 2.1. □

Proof of Proposition 4.1. Let H_n denote the distribution function of $\|\hat{\mathcal{M}}'_n \mathbb{G}_n^*\|$ conditional on $\mathcal{X}_n$, and let H denote the distribution function of $\|\hat{\mathbb{G}}\|$. Since the norm $\|\cdot\|$ is Lipschitz continuous we may apply Theorem 3.2 with $d_{\mathcal{P}}$ the bounded Lipschitz metric to obtain $E(f(\|\hat{\mathcal{M}}'_n \mathbb{G}_n^*\|) \,|\, \mathcal{X}_n) \to Ef(\|\hat{\mathbb{G}}\|)$ in probability for all Lipschitz continuous maps $f : \mathbf{R} \to [0,1]$. It now follows from Lemma 10.11(i) of Kosorok (2008b) that we have $H_n(c) \to H(c)$ in probability for all continuity points c of H. Fix $\epsilon > 0$ and choose two such continuity points c_1 and c_2 satisfying $q^{1-\alpha} - \epsilon < c_1 < q^{1-\alpha} < c_2 < q^{1-\alpha} + \epsilon$ and $H(c_1) + \delta < 1 - \alpha < H(c_2) - \delta$ for some $\delta > 0$; this is always possible since H is strictly increasing at $q^{1-\alpha}$. Observe that if $|q_n^{1-\alpha} - q^{1-\alpha}| > \epsilon$ then either $1-\alpha \leq H_n(c_1)$ or $H_n(c_2) \leq 1 - \alpha$. Thus,

$$\begin{aligned} P(|q_n^{1-\alpha} - q^{1-\alpha}| > \epsilon) &\leq P(1 - \alpha \leq H_n(c_1)) + P(H_n(c_2) \leq 1 - \alpha) \\ &\leq P(H_n(c_1) \geq H(c_1) + \delta) + P(H_n(c_2) \leq H(c_2) - \delta). \end{aligned}$$

Since $H_n(c_1) \to H(c_1)$ and $H_n(c_2) \to H(c_2)$ in probability, it follows that $q_n^{1-\alpha} \to q^{1-\alpha}$ in probability, as claimed. Convergence of the coverage probability to $1 - \alpha$ now follows from Theorem 2.1, the continuous mapping theorem, and the fact that H is continuous at $q^{1-\alpha}$. □

Proof of Proposition 4.2. It may be shown that $\bar{q}_n^{1-\alpha} \to \bar{q}^{1-\alpha}$ in probability and that $P(\|\mathbb{F}_n - F\| \leq n^{-1/2}\bar{q}_n^{1-\alpha}) \to 1-\alpha$ by arguing as in the proof of Proposition 4.1, but using Assumption 3.1 in place of Theorem 3.2. Weak contractivity of

$\mathcal{M}$ implies that $\|\hat{\mathbb{F}}_n - F\| \le \|\mathbb{F}_n - F\|$, so we have $P(\|\hat{\mathbb{F}}_n - F\| \le n^{-1/2}\bar{q}_n^{1-\alpha}) \ge P(\|\mathbb{F}_n - F\| \le n^{-1/2}\bar{q}_n^{1-\alpha})$, and the desired result follows by taking the limit inferior of both sides. □

Proof of Theorem 4.1. We first show that

$$d_{\mathcal{P}}(\mathcal{L}(\sqrt{n}(\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n) \mid \mathcal{X}_n), \mathcal{L}(\mathbb{G})) \to 0 \tag{5.8}$$

almost surely (a.s.), with $d_{\mathcal{P}}$ the bounded Lipschitz metric. Let P_n be the probability measure corresponding to $\hat{\mathbb{F}}_n$, and let $\mathbb{P}_n$ be the probability measure corresponding to $\mathbb{F}_n^\dagger$, the empirical distribution of n iid bootstrap draws from P_n. Let $\mathbb{G}_{n,P_n} = \sqrt{n}(\mathbb{P}_n - P_n)$. Then we may write

$$\sqrt{n}(\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n)(x) = \mathbb{G}_{n,P_n} f_x \tag{5.9}$$

where $f_x = \mathbb{1}_{(-\infty,x]}$ for all $x \in \mathbf{R}$. We aim to invoke Lemma 2.8.7 of van der Vaart and Wellner (1996). First, $\mathcal{F} = \{f_x : x \in \mathbf{R}\}$ is Donsker and pre-Gaussian uniformly in all probability measures on $\mathbf{R}$ (Giné and Zinn, 1991), and hence uniformly in $\{P_n\}$. Condition 2.8.6 of Lemma 2.8.7 of van der Vaart and Wellner (1996) is trivially satisfied since $\mathcal{F}$ is uniformly bounded. It remains for us to verify that their condition 2.8.5 holds a.s. Let P_0 be the probability measure corresponding to F. Pick $f_x, f_y \in \mathcal{F}$ and assume $x < y$ without loss of generality. Then using the notation in van der Vaart and Wellner (1996),

$$|\rho_{P_n}(f_x, f_y) - \rho_{P_0}(f_x, f_y)| = |\sqrt{P_n(x,y] - (P_n(x,y])^2} - \sqrt{P_0(x,y] - (P_0(x,y])^2}|.$$

Condition 2.8.5 of van der Vaart and Wellner (1996) requires that $\rho_{P_n}(f_x, f_y) \to \rho_{P_0}(f_x, f_y)$ uniformly in x, y. By applying Marshall's lemma and the Glivenko-Cantelli theorem we see that $P_n(x,y] \to P_0(x,y]$ uniformly in x, y a.s., and so condition 2.8.5 holds a.s. Thus Lemma 2.8.7 of van der Vaart and Wellner (1996) implies that $\mathbb{G}_{n,P_n} \rightsquigarrow \mathbb{G}_{P_0}$ in $\ell^\infty(\mathcal{F})$ a.s., where $\mathbb{G}_{P_0}$ is tight, centered and Gaussian with $E\mathbb{G}_{P_0} f_x \mathbb{G}_{P_0} f_y = P_0 f_x f_y - P_0 f_x P_0 f_y$. In view of (5.9) and the fact that $E\mathbb{G}(x)\mathbb{G}(y) = E\mathbb{G}_{P_0} f_x \mathbb{G}_{P_0} f_y$, it follows that (5.8) holds a.s.

With the conditional weak convergence (5.8) in hand, we may apply Theorem 2.2 of Kosorok (2008a) to obtain

$$(\sqrt{n}(\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n), \sqrt{n}(\mathbb{F}_n - F)) \rightsquigarrow (\mathbb{G}, \mathbb{G}') \quad \text{in } \ell^\infty \times \ell^\infty. \tag{5.10}$$

(Note that (5.8) holding a.s. implies $\sqrt{n}(\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n) \overset{\mathrm{P}}{\underset{W}{\rightsquigarrow}} \mathbb{G}$ in Kosorok's notation, due to the fact that $\sqrt{n}(\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n)$ is ball measurable and càdlàg; see Bücher and Ruppert, 2013, p. 224.) For each n let $g_n : \ell^\infty \times \ell^\infty \to \ell^\infty \times \ell^\infty$ be the map

$$g_n(\theta_1, \theta_2) = (\theta_1, \sqrt{n}(\mathcal{M}(F + n^{-1/2}\theta_2) - \mathcal{M}(F))),$$

and note that for any sequence of pairs $(\theta_{1,n}, \theta_{2,n})$ in $\ell^\infty \times \ell^\infty$ converging to some $(\theta_1, \theta_2) \in \ell^\infty \times C_0$, Proposition 2.1 ensures that $g_n(\theta_{1,n}, \theta_{2,n}) \to (\theta_1, \mathcal{M}'_F(\theta_2))$. We may therefore apply the extended continuous mapping theorem with the maps g_n to the weak convergence (5.10), yielding

$$(\sqrt{n}(\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n), \sqrt{n}(\hat{\mathbb{F}}_n - F)) \rightsquigarrow (\mathbb{G}, \mathcal{M}'_F(\mathbb{G}')) \quad \text{in } \ell^\infty \times \ell^\infty.$$

Applying the continuous mapping theorem with the map $(\theta_1, \theta_2) \mapsto (\theta_1 + \theta_2, \theta_2)$ then gives

$$\sqrt{n}((\mathbb{F}_n^\dagger, \hat{\mathbb{F}}_n) - (F, F)) \rightsquigarrow (\mathbb{G} + \mathcal{M}_F'(\mathbb{G}'), \mathcal{M}_F'(\mathbb{G}')) \quad \text{in } \ell^\infty \times \ell^\infty.$$

Finally, we apply the functional Delta method with the map $(\theta_1, \theta_2) \mapsto \mathcal{M}(\theta_1) - \theta_2$ to obtain

$$\sqrt{n}(\mathcal{M}\mathbb{F}_n^\dagger - \hat{\mathbb{F}}_n) \rightsquigarrow \mathcal{M}_F'\left(\mathbb{G} + \mathcal{M}_F'(\mathbb{G}')\right) - \mathcal{M}_F'(\mathbb{G}') \quad \text{in } \ell^\infty,$$

as claimed. □

Appendix A: Extension to isotonic regression

While the primary focus of our paper has been on the behavior of estimators of concave distribution functions, the approach we have taken may be used to study other shape constrained estimators. In this appendix we briefly explain how the results on the Hadamard directional differentiability of the LCM operator developed here and in Beare and Moon (2015) may be used to characterize the limiting behavior of the greatest convex minorant (GCM) of the cumulated sum diagram (CSD), whose left-derivative is the isotonic regression estimator.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n iid pairs of random variables satisfying

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $m : \mathbf{R} \to \mathbf{R}$ is nondecreasing and $\varepsilon_1, \ldots, \varepsilon_n$ are iid centered random variables independent of $X_1, \ldots, X_n$. We impose the following technical conditions.

Assumption A.1. (a) The second moment $\sigma^2 = E\varepsilon_1^2$ is positive and finite; (b) the distribution function F of X_1 has compact support $[a, b]$ and is continuously differentiable on that support with strictly positive derivative f; (c) m is continuously differentiable on $[a, b]$ with derivative m'.

Let the X_i's arranged in ascending order be denoted by $X_{(1)}, \ldots, X_{(n)}$, and the corresponding Y_i's and ε_i's by $Y_{(1)}, \ldots, Y_{(n)}$ and $\varepsilon_{(1)}, \ldots, \varepsilon_{(n)}$. The isotonic regression estimator of m proposed by Brunk (1958) is given by

$$\hat{m}_n(X_{(i)}) = \max_{s \le i} \min_{t \ge i} \sum_{j=s}^{t} \frac{Y_{(j)}}{t - s + 1}$$

at each observation $X_{(i)}$. Alternatively, $\hat{m}_n(X_{(i)})$ is given by the left-derivative at i of the GCM over $[0, n]$ of the CSD of $Y_{(1)}, \ldots, Y_{(n)}$ (Brunk, 1958; Mukerjee, 1988). Recall that this CSD is given by $\{(k, S_n(k)) : k = 0, \ldots, n\}$ where $S_n(0) = 0$ and

$$S_n(k) = \sum_{j=1}^{k} Y_{(j)}, \quad k = 1, \ldots, n.$$

To apply our theory we reformulate the CSD as a random element of $\ell^\infty([0,1])$ by setting

$$\mathbb{S}_n(u) = \frac{1}{n}\sum_{i=0}^{[nu]} Y_{(i)} + \frac{nu - [nu]}{n} Y_{([nu]+1)} \tag{A.1}$$

for $u \in [0,1]$, where we define $Y_{(0)} = Y_{(n+1)} = 0$ and for $x \in \mathbf{R}$ we denote by $[x]$ the largest integer in $[x-1, x]$. The second term on the right-hand side of (A.1) is bounded in absolute value by $n^{-1}\max_{i\le n}|Y_i|$, which is $o_P(n^{-1/2})$ since the Y_i's are iid and $EY_1^2 < \infty$ under Assumption A.1. The first term admits the decomposition

$$\frac{1}{n}\sum_{i=0}^{[nu]} Y_{(i)} = \frac{1}{n}\sum_{i=0}^{[nu]} m(X_{(i)}) + \frac{1}{n}\sum_{i=0}^{[nu]} \varepsilon_{(i)}. \tag{A.2}$$

The ε_i's are iid and independent of the X_i's, so the second term on the right-hand side of (A.2) is independent of the first and is equal in law to $n^{-1}\sum_{i=1}^{[nu]}\varepsilon_i$, which satisfies

$$n^{-1}\sum_{i=1}^{[nu]} \varepsilon_i \rightsquigarrow \sigma\mathbb{W}(u)$$

under Assumption A.1(a), where $\mathbb{W}$ is a standard Brownian motion; throughout this appendix, $\rightsquigarrow$ signifies weak convergence in $\ell^\infty([0,1])$ unless otherwise stated. The first term on the right-hand side of (A.2) can be rewritten as

$$\frac{1}{n}\sum_{i=0}^{[nu]} m(X_{(i)}) = \int_0^{[nu]/n} m(\mathbb{Q}_n(t))\mathrm{d}t = \int_0^u m(\mathbb{Q}_n(t))\mathrm{d}t + O_P(n^{-1}),$$

where $\mathbb{Q}_n$ is the empirical quantile function for $X_1, \ldots, X_n$. Under Assumption A.1(b) we may apply the functional Delta method in conjunction with Lemma 21.4(ii) of van der Vaart (1998) to obtain $\sqrt{n}(\mathbb{Q}_n - Q) \rightsquigarrow -\mathbb{B}/(f(Q))$ in $\ell^\infty((0,1))$, where $\mathbb{B}$ is a standard Brownian bridge, and Q is the quantile function for X_1. Another application of the functional Delta method, but this time in conjunction with an obvious generalization of Lemma 22.9 of van der Vaart (1998) and appealing to Assumption A.1(c), can be used to show that

$$\sqrt{n}\left(\int_0^u m(\mathbb{Q}_n(t))\mathrm{d}t - \int_0^u m(Q(t))\mathrm{d}t\right) \rightsquigarrow -\int_0^u \frac{m'(Q(t))}{f(Q(t))}\mathbb{B}(t)\mathrm{d}t.$$

Let $S(u) = \int_0^u m(Q(t))\mathrm{d}t$. We have proved the following result.

Proposition A.1. *Under Assumption A.1 we have*

$$\sqrt{n}\,(\mathbb{S}_n(u) - S(u)) \rightsquigarrow -\int_0^u \frac{m'(Q(t))}{f(Q(t))}\mathbb{B}(t)\mathrm{d}t + \sigma\mathbb{W}(u) \quad \text{in } \ell^\infty([0,1]),$$

where $\mathbb{B}$ is a standard Brownian bridge, $\mathbb{W}$ is a standard Brownian motion, and $\mathbb{B}$ and $\mathbb{W}$ are independent.

Let $\hat{\mathbb{S}}_n = -\mathcal{M}_{[0,1]}(-\mathbb{S}_n)$, the GCM of $\mathbb{S}_n$. The isotonic regression estimate $\hat{m}_n(X_{(i)})$ is equal to the left-derivative of $\hat{\mathbb{S}}_n$ at i/n. We can obtain a characterization of the limiting behavior of $\hat{\mathbb{S}}_n$ from Proposition A.1 by applying Lemma 3.2 of Beare and Moon (2015), which is similar to Proposition 2.1 above but applies to LCMs of functions in $\ell^\infty([0,1])$. This result requires the point in $\ell^\infty([0,1])$ at which we differentiate to be concave, continuously differentiable, and weakly increasing from zero to one. To apply it we therefore set

$$\lambda = -\int_0^1 (m(Q(t)) - m(b))\,\mathrm{d}t \geq 0,$$

and define

$$S_0(u) = \begin{cases} \frac{1}{\lambda}\int_0^u (m(Q(t)) - m(b))\,\mathrm{d}t & \text{if } \lambda > 0 \\ -u & \text{if } \lambda = 0, \end{cases}$$

so that $S(u) = \lambda S_0(u) + L_0(u)$, where $L_0(u) = m(b)u$. Note that $-S_0$ is concave, continuously differentiable, and weakly increasing from zero to one.

Suppose that $\lambda > 0$. The weak convergence established in Proposition A.1 may then be rewritten as

$$\sqrt{n}\left(\frac{L_0(u) - \mathbb{S}_n(u)}{\lambda} - (-S_0(u))\right) \rightsquigarrow \frac{1}{\lambda}\int_0^u \frac{m'(Q(t))}{f(Q(t))}\mathbb{B}(t)\mathrm{d}t - \frac{\sigma}{\lambda}\mathbb{W}(u). \quad \text{(A.3)}$$

Since L_0 is affine and $\lambda > 0$, Lemma 2.1 of Durot and Tocquet (2003) implies that the LCM of $\lambda^{-1}(L_0(u) - \mathbb{S}_n(u))$ is $\lambda^{-1}(L_0(u) - \hat{\mathbb{S}}_n(u))$. Therefore, in view of the weak convergence (A.3) and the Hadamard directional differentiability of $\mathcal{M}_{[0,1]}$ established in Lemma 3.2 of Beare and Moon (2015), an application of the functional Delta method yields

$$\sqrt{n}\left(\frac{L_0 - \hat{\mathbb{S}}_n}{\lambda} - (-S_0)\right) \rightsquigarrow \mathcal{M}'_{[0,1],-S_0}\left(\frac{1}{\lambda}\int_0^\cdot \frac{m'(Q(t))}{f(Q(t))}\mathbb{B}(t)\mathrm{d}t - \frac{\sigma}{\lambda}\mathbb{W}(\cdot)\right),$$

which in view of the positive homogeneity of $\mathcal{M}'_{[0,1],-S_0}$ may be rewritten as

$$\sqrt{n}(\hat{\mathbb{S}}_n - S) \rightsquigarrow -\mathcal{M}'_{[0,1],-S_0}\left(\int_0^\cdot \frac{m'(Q(t))}{f(Q(t))}\mathbb{B}(t)\mathrm{d}t - \sigma\mathbb{W}(\cdot)\right). \quad \text{(A.4)}$$

The directional derivative $\mathcal{M}'_{[0,1],-S_0}$ in direction $h \in C([0,1])$ majorizes h by concave functions on regions over which $-S_0$ is affine but acts like an identity map elsewhere, similar to the derivative in Proposition 2.1. In particular, if m is strictly increasing on the support of F then $-S_0$ is strictly concave, and so

$$\sqrt{n}(\hat{\mathbb{S}}_n(u) - S(u)) \rightsquigarrow -\int_0^u \frac{m'(Q(t))}{f(Q(t))}\mathbb{B}(t)\mathrm{d}t + \sigma\mathbb{W}(u),$$

indicating asymptotic equivalence of $\hat{\mathbb{S}}_n$ and $\mathbb{S}_n$.

Suppose instead that $\lambda = 0$, so that $m' = 0$ on the support of F. In this case the weak convergence established in Proposition A.1 may be rewritten as

$$\sqrt{n}\,(\mathbb{S}_n - S) \rightsquigarrow \sigma\mathbb{W}, \tag{A.5}$$

with $S = L_0$. Since L_0 is affine, Lemma 2.1 of Durot and Tocquet (2003) implies that the LCM of $\sqrt{n}(L_0 - \mathbb{S}_n)$ is $\sqrt{n}(L_0 - \hat{\mathbb{S}}_n)$. From (A.5) and the continuous mapping theorem we therefore have $\sqrt{n}(\hat{\mathbb{S}}_n - S) \rightsquigarrow -\mathcal{M}_{[0,1]}(-\sigma\mathbb{W})$. Lemma 3.2 of Beare and Moon (2015) shows that $\mathcal{M}'_{[0,1],-S_0} = \mathcal{M}_{[0,1]}$ when $\lambda = 0$, and so we have established that the weak convergence (A.4) holds also when $\lambda = 0$. Thus we have proved the following result on the limiting behavior of $\hat{\mathbb{S}}_n$.

Proposition A.2. *Under Assumption A.1 we have the weak convergence* (A.4) *in* $\ell^\infty([0,1])$*, where* $\mathbb{B}$ *is a standard Brownian bridge,* $\mathbb{W}$ *is a standard Brownian motion, and* $\mathbb{B}$ *and* $\mathbb{W}$ *are independent.*

Our results on the limiting behavior of $\mathbb{S}_n$ and $\hat{\mathbb{S}}_n$ suggest a simple test of the null hypothesis that the nondecreasing function m is in fact flat. Under this null, it may be shown using the weak convergence (A.5), Lemma 2.1 of Durot and Tocquet (2003) and the continuous mapping theorem that

$$\sqrt{n}(\hat{\mathbb{S}}_n - \mathbb{S}_n) \rightsquigarrow \sigma\,(\mathcal{M}\mathbb{W} - \mathbb{W})\,.$$

On the other hand, if m is strictly increasing then $\mathbb{S}_n$ and $\hat{\mathbb{S}}_n$ are asymptotically equivalent, and we have $\sqrt{n}\|\hat{\mathbb{S}}_n - \mathbb{S}_n\| \to 0$ in probability. This suggests using a test statistic of the form $T_n = \sqrt{n}\|\hat{\mathbb{S}}_n - \mathbb{S}_n\|/\hat{\sigma}_n$, where $\hat{\sigma}_n$ is a consistent estimator of σ, and rejecting the null of flatness when T_n is smaller than the α-quantile of $\|\mathcal{M}\mathbb{W} - \mathbb{W}\|$. Such a procedure will reject with probability approaching α when m is flat, and with probability approaching one when m is strictly increasing.

References

Balabdaoui, F., Jankowski, H., Pavlides, M., Seregin, A. and Wellner, J. A. (2011). On the Grenander estimator at zero. *Statistica Sinica*, 21(2):873–899. MR2829859

Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica*, 78(1):395–410. MR2642867

Beare, B. K. (2012). Archimedean copulas and temporal dependence. *Econometric Theory*, 28(6):1165–1185. MR2998469

Beare, B. K. and Moon, J. -M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory*, 31(3):471–492. MR3348455

Beare, B. K. and Schmidt, L. D. W. (2016). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics*, 31(2):338–356. MR3481366

Beare, B. K. and Seo, J. (2014). Time irreversible copula-based Markov models. *Econometric Theory*, 30(5):923–960. MR3252030

Beran, R. and Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics*, 13(1):95–115. MR0773155

Beran, R. and Srivastava, M. S. (1987). Correction: Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics*, 15(1):470–471. MR0885755

Bickel, P. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7(1):1–31. MR1441142

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionelles. *Annales de l'Institut Henri Poincaré* (B), 19(3):281–296. MR0725561

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

Bücher, A. and Ruppert, M. (2013). Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *Journal of Multivariate Analysis*, 116:208–229. MR3049901

Carolan, C. A. (2002). The least concave majorant of the empirical distribution function. *Canadian Journal of Statistics*, 30(2):317–328. MR1926068

Carolan, C. A. and Dykstra, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canadian Journal of Statistics*, 27(3):557–566. MR1745821

Carolan, C. A. and Dykstra, R. (2001). Marginal densities of the least concave majorant of Brownian motion. *Annals of Statistics*, 29(6):1732–1750. MR1891744

Carolan, C. A. and Tebbs, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two sample problem. *Biometrika*, 92(1):159–171. MR2158617

Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In Dehling, H., Mikosch, T. and Sørensen, M. (Eds.), *Empirical Process Techniques for Dependent Data*, 3–113. Birkhäuser. MR1958777

Delgado, M. A. and Escanciano, J. C. (2012). Distribution-free tests of stochastic monotonicity. *Journal of Econometrics*, 170(1):68–75. MR2955940

Delgado, M. A. and Escanciano, J. C. (2013). Conditional stochastic dominance testing. *Journal of Business and Economic Statistics*, 31(1):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(1):125–140. MR1207311

Durot, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Annals of Statistics*, 35(3):1080–1104. MR2341699

Durot, C., Kulikov, V. N. and Lopuhaä, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Annals of Statistics*, 40(3):1578–1608. MR3015036

Durot, C. and Tocquet, A. -S. (2003). On the distance between the empir-

ical process and its concave majorant in a monotone regression framework. *Annales de l'Institut Henri Poincaré* (B), 39(2):217–240. MR1962134

Eggermont, P. P. B. and LaRiccia, V. N. (2000). Maximum likelihood estimation of smooth monotone and unimodal densities. *Annals of Statistics*, 28(3):922–947. MR1792794

Eggermont, P. P. B. and LaRiccia, V. N. (2001). *Maximum Penalized Likelihood Estimation, Volume I: Density Estimation*. Springer Series in Statistics, Springer, New York. MR1837879

Fang, Z. and Santos, A. (2015). Inference on directionally differentiable functions. arXiv:1404.3763v2 MR3407344

Giné, E. and Zinn, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Annals of Probability*, 19(2):758–782. MR1106285

Grenander, U. (1956). On the theory of mortality measurement. Part II. *Scandinavian Actuarial Journal*, 39:125–153. MR0093415

Groeneboom, P. (1985). Estimating a monotone density. In Le Cam, L. M. and Olshen, R. A. (Eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. 2, 539–555. Wadsworth Statistics/Probability Series, Wadsworth, Belmont. MR0822052

Groeneboom, P., Hooghiemstra, G. and Lopuhaä, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Annals of Statistics*, 27(4):1316–1347. MR1740109

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints:Estimators, Algorithms and Asymptotics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, New York. MR3445293

Jankowski, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Annals of Statistics*, 42(2):625–653. MR3210981

Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Probability Theory and Related Fields*, 34(1):73–85. MR0397974

Kosorok, M. R. (2008a). Bootstrapping the Grenander estimator. In Balakrishnan, N., Peña, E. A. and Silvapulle, M. J. (Eds.), *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, 282–292. Institute of Mathematical Statistics, Beachwood. MR2462212

Kosorok, M. R. (2008b). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics, Springer, New York. MR2724368

Kulikov, V. N. and Lopuhaä, H. P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *Annals of Statistics*, 33(5):2228–2255. MR2211085

Kulikov, V. N. and Lopuhaä, H. P. (2006a). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Annals of Statistics*, 34(2):742–768. MR2283391

Kulikov, V. N. and Lopuhaä, H. P. (2006b). The limit process of the difference between the empirical distribution function and its concave majorant. *Statistics and Probability Letters*, 76(16):1781–1786. MR2274141

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(2):356–377. MR2391249

Mukerjee, H. (1988). Monotone nonparametric regression. *Annals of Statistics*, 16(2):741–750. MR0947574

Paparoditis, E. and Politis, D. N. (2002). The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference*, 108(1):301–328. MR1947405

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Series in Statistics, Springer, New York. MR0762984

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Series A*, 31:23–36. MR0267677

Radulović, D. (2002). On the bootstrap and empirical processes for dependent sequences. In Dehling, H., Mikosch, T. and Sørensen, M. (Eds.), *Empirical Process Techniques for Dependent Data*, 345–364. Birkhäuser, Boston. MR1958789

Sen, B., Banerjee, M. and Woodroofe, M. (2010). Inconsistency of bootstrap: the Grenander estimator. *Annals of Statistics*, 38(4):1953–1977. MR2676880

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487. MR1080259

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30(1):169–186. MR1118896

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics, Wiley, New York. MR0838963

van der Vaart, A. (1994). Weak convergence of smoothed empirical processes. *Scandinavian Journal of Statistics*, 21(4):501–504. MR1310093

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge. MR1652247

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics, Springer, New York. MR1385671

Volgushev, S. and Shao, X. (2014). A general approach to the joint asymptotic analysis of statistics from subsamples. *Electronic Journal of Statistics*, 8(1):390–431. MR3195121

Walther, G. (2009) Inference and modeling with log-concave distributions. *Statistical Science*, 24(3):319–327. MR2757433

Wang, Y. (1994). The limit distribution of the concave majorant of an empirical distribution function. *Statistics and Probability Letters*, 20(1):81–84. MR1294808

Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statistica Sinica*, 3(2):501–515. MR1243398