BOOTSTRAP-ASSISTED INFERENCE
FOR GENERALIZED GRENANDER-TYPE ESTIMATORS

By Matias D. Cattaneo[1,a], Michael Jansson[2,b] and Kenichi Nagasawa[3,c]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*cattaneo@princeton.edu*
[2]*Department of Economics, University of California, Berkeley,* [b]*mjansson@econ.berkeley.edu*
[3]*Department of Economics, University of Warwick,* [c]*kenichi.nagasawa@warwick.ac.uk*

Westling and Carone (*Ann. Statist.* **48** (2020) 1001–1024) proposed a framework for studying the large sample distributional properties of generalized Grenander-type estimators, a versatile class of nonparametric estimators of monotone functions. The limiting distribution of those estimators is representable as the left derivative of the greatest convex minorant of a Gaussian process whose monomial mean can be of unknown order (when the degree of flatness of the function of interest is unknown). The standard nonparametric bootstrap is unable to consistently approximate the large sample distribution of the generalized Grenander-type estimators even if the monomial order of the mean is known, making statistical inference a challenging endeavour in applications. To address this inferential problem, we present a bootstrap-assisted inference procedure for generalized Grenander-type estimators. The procedure relies on a carefully crafted, yet automatic, transformation of the estimator. Moreover, our proposed method can be made "flatness robust" in the sense that it can be made adaptive to the (possibly unknown) degree of flatness of the function of interest. The method requires only the consistent estimation of a single scalar quantity, for which we propose an automatic procedure based on numerical derivative estimation and the generalized jackknife. Under random sampling, our inference method can be implemented using a computationally attractive exchangeable bootstrap procedure. We illustrate our methods with examples and we also provide a small simulation study. The development of formal results is made possible by some technical results that may be of independent interest.

1. Introduction. Monotone function estimators have received renewed attention in statistics, biostatistics, econometrics, machine learning, and other data science disciplines. See Groeneboom and Jongbloed (2014, 2018) for a textbook introduction and a review article, respectively, the latter being published in a special issue devoted to nonparametric inference under shape constraints. More recently, Westling and Carone (2020) expanded the scope and applicability of monotone function estimators by embedding many such estimators in a unified framework of generalized Grenander-type estimators. Estimation problems covered by Westling and Carone's (2020) general theory include many practically relevant examples such as monotone density, regression and hazard estimation, possibly with censoring and/or covariate adjustment.

The large sample theory developed by Westling and Carone (2020) offers a general distributional approximation involving the left derivative of the Greatest Convex Minorant (GCM) of a Gaussian process whose mean and covariance kernel depend on unknown functions. Furthermore, both the convergence rate of the estimator and the shape of the mean appearing in the representation of the limiting distribution depend on whether the unknown function of

interest exhibits certain degeneracies. For these reasons, the large sample distributional approximation for generalized Grenander-type estimators can be difficult to employ in practice for inference purposes. In their concluding remarks, Westling and Carone (2020) recognize these limitations and pose the question of whether it would be possible to employ bootstrap-assisted methods to conduct automatic/robust statistical inference within their framework.

As is well documented, the standard nonparametric bootstrap does not provide a valid distributional approximation for the generalized Grenander-type estimators (Kosorok (2008), Sen, Banerjee and Woodroofe (2010)). This fact has led scholars to rely on other bootstrap schemes such as subsampling (Politis and Romano (1994)), the smoothed bootstrap (Kosorok (2008), Sen, Banerjee and Woodroofe (2010)), the m-out-of-n bootstrap (Sen, Banerjee and Woodroofe (2010), Lee and Yang (2020)), or the numerical bootstrap (Hong and Li (2020)). See also Cavaliere and Georgiev (2020), and references therein, for some related recent results. Those approaches could in principle be used to construct bootstrap-based inference methods for (some members of the class of) generalized Grenander-type estimators, but they all would require employing specific regularized multidimensional bootstrap distributions or related quantities involving multiple smoothing and tuning parameters, rendering those approaches potentially difficult to implement in practice. Furthermore, those methods would not be robust to unknown degeneracies determining the convergence rate and shape of the limiting distribution without additional modifications. For example, subsampling methods require knowledge of the precise convergence rate of the statistic, or estimation thereof, as a preliminary step (Politis, Romano and Wolf (1999)). Another resampling approach was recently proposed by Kuchibhotla, Balakrishnan and Wasserman (2024) and Mallick, Sarkar and Kuchibhotla (2024), which offers valid confidence intervals under an approximate symmetry condition of the limiting distribution of the estimator.

We complement existing methods by introducing a novel bootstrap-assisted inference approach that restores validity of bootstrap methods by reshaping of the ingredients of the generalized Grenander-type estimator. Our approach is motivated by a constructive interpretation of the source of the bootstrap failure. As a by-product, our interpretation explicitly isolates the role of unknown degeneracies determining the precise form of the limiting distribution, a feature of the interpretation which allow us to develop an automatic inference method that is robust to such degeneracies, ultimately resulting in a more robust bootstrap-assisted inference approach. In the case of random sampling, we show that our method can be implemented using a computationally attractive exchangeable bootstrap procedure. For completeness, we also discuss implementation issues, offering a fully automatic (i.e., data-driven) valid inference method for generalized Grenander-type estimators. Some of the ideas underlying our approach are similar to ideas used in Cattaneo, Jansson and Nagasawa (2020), where we introduced a bootstrap-based distributional approximation for M-estimators with possibly nonstandard Chernoff (1964)-type asymptotic distributions (Kim and Pollard (1990), Seo and Otsu (2018)). Generalized Grenander-type estimators are not M-estimators, however, and as further explained in the next paragraph the analysis of generalized Grenander-type estimators turns out to necessitate the development of technical tools that play no role in the analysis of M-estimators.

Although valid distributional approximations for monotone estimators can be obtained in a variety of ways (Groeneboom and Jongbloed (2014)), by far the most common approach is to employ the *switch relation* (Groeneboom (1985)) to re-express the cumulative distribution function (cdf) of the suitably normalized monotone estimator in terms of a probability statement about the maximizer of a certain stochastic process whose large sample properties can in turn be analyzed by employing standard empirical process methods (van der Vaart and Wellner (1996)). From a technical perspective, this approach requires (at least) two ingredients in order to be successful, namely results establishing (i) validity of the switch relation

and (ii) continuity of the limiting cdf of the maximizer of a stochastic process. In the process of developing our main results, we shed new light on both (i) and (ii). First, we show by example that Westling and Carone's ((2020), Supplement) generalization of the switch relation is incomplete as stated and then propose a modification. Our modification of Westling and Carone's ((2020), Supplement) Lemma 1 shows that the conclusion of that lemma is valid once an additional assumption is made. The additional assumption seems very mild, being satisfied in all examples considered by Westling and Carone (2020) and all other examples of which we are aware, including some new examples we consider in this paper. Second, we present a lemma establishing continuity of the cdf of the maximizer of a Gaussian process, a result which can in turn be used to establish continuity of the cdf of the suitably normalized generalized Grenander-type estimator. Interestingly, although these continuity properties are important when deriving limiting distributions with the help of the switch relation and justifying bootstrap-type inference procedures, respectively, it would appear that explicit statements of them are unavailable in the existing literature. (A prominent exception is the one where the limiting distribution is a scaled Chernoff distribution, which is known to be absolutely continuous.) For further details on (i) and (ii), see the Appendix.

In the remainder of this introductory section we outline key notation and definitions used throughout the paper. Section 2 then recalls the setup of Westling and Carone (2020) and presents a version of their main distributional approximation result for generalized Grenander-type estimators. Section 3 contains our main results about bootstrap-assisted distributional approximations, while Section 4 discusses implementation issues, including tuning parameter selection and a computationally attractive weighted bootstrap procedure. Section 5 illustrates our general theory by means of prominent examples, while Section 6 reports numerical results from a small-scale simulation experiment. The Appendix reports the two technical results alluded to in the previous paragraph. All proofs and other technical details are given in the Supplementary Material (Cattaneo, Jansson and Nagasawa (2024)).

1.1. *Notation and definitions.* For an interval $I \subseteq \mathbb{R}$ and a function $f : I \to \mathbb{R}$, $\mathsf{GCM}_I(f)$ denotes its greatest convex minorant (on I) and if f is nondecreasing and right-continuous, then f^- denotes its generalized inverse; that is, $f^-(x) = \inf\{u \in I : f(u) \geq x\}$, where the dependence of f^- on I has been suppressed (and where the infimum of the empty set is $\sup I$). Also, if f is nondecreasing and right-continuous, we write $f(x-) = \lim_{\eta \downarrow 0} f(x - \eta)$ and $f^-(y+) = \lim_{\eta \downarrow 0} f^-(y + \eta)$. Assuming the relevant derivatives exist, ∂^q denotes the qth partial derivative (operator) and ∂_- denotes the left derivative (operator). In addition, $f \circ g$ denotes the composition of f and g; that is, $(f \circ g)(x) = f(g(x))$. Finally, for $a, b \in \mathbb{R}$, we write $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, and $a^+ = \max(a, 0)$, and $\lfloor a \rfloor$ denotes the integer part of a.

Limits are taken as $n \to \infty$, unless otherwise stated. For two (possibly) random sequences $\{a_n\}$ and $\{b_n\}$, $a_n = O_{\mathbb{P}}(b_n)$ is shorthand for $\limsup_{\epsilon \to \infty} \limsup_n \mathbb{P}[|a_n/b_n| \geq \epsilon] = 0$, $a_n = o_{\mathbb{P}}(b_n)$ is shorthand for $\limsup_{\epsilon \to 0} \limsup_n \mathbb{P}[|a_n/b_n| \geq \epsilon] = 0$, and the subscript "$\mathbb{P}$" on "$O$" and "$o$" is often omitted when $\{a_n\}$ and $\{b_n\}$ are nonrandom. We use $\to_{\mathbb{P}}$ to denote convergence in probability and $\rightsquigarrow$ to denote weak convergence, where, for a stochastic process indexed by $\mathbb{R}$, convergence is in the topology of uniform convergence on compacta. When analyzing the bootstrap, $\mathbb{P}_n^*$ denotes the probability measure under the bootstrap distribution conditional on the original data and $\rightsquigarrow_{\mathbb{P}}$ denotes weak convergence in probability conditionally on the original data. For more details, see van der Vaart and Wellner (1996).

2. Setup. Our setup is that of Westling and Carone (2020). The goal is to conduct inference on $\theta_0(\mathsf{x})$, where, for some interval $I \subseteq \mathbb{R}$, θ_0 is nondecreasing on I and x is an interior point of I. Assuming it is well defined, the function Θ_0 given by

$$\Theta_0(x) = \int_{\inf I}^x \theta_0(v)\, dv$$

is convex on I and therefore enjoys the property that if θ_0 is left-continuous at x, then

$$\theta_0(\mathsf{x}) = \partial_- \mathsf{GCM}_I(\Theta_0)(\mathsf{x}).$$

An estimator of $\theta_0(\mathsf{x})$ obtained by replacing Θ_0 and I in the preceding display with estimators is said to be of the Grenander-type, a canonical example of this class of estimators being the celebrated Grenander (1956) estimator of a nondecreasing density.

EXAMPLE 1 (Monotone Density Estimation). Suppose $X_1, \ldots, X_n$ are i.i.d. copies of a continuously distributed nonnegative random variable X whose density f_0 is nondecreasing on its support $[0, u_0]$, where u_0 is (possibly) unknown. For $\mathsf{x} \in (0, u_0)$, the Grenander (1956) estimator of $f_0(\mathsf{x})$ is

$$\widehat{f}_n(\mathsf{x}) = \partial_- \mathsf{GCM}_{[0, \widehat{u}_n]}(\widehat{F}_n)(\mathsf{x}),$$

where $\widehat{u}_n = \max_{1 \leq i \leq n} X_i \vee \mathsf{x}$ and where $\widehat{F}_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$ is the empirical cdf. Section 5.1 and the Supplementary Material discuss more general monotone density estimators allowing for censoring and/or covariate adjustment (e.g., van der Laan and Robins (2003)).

To define the class of generalized Grenander-type estimators, let $\psi_0 = \theta_0 \circ \Phi_0^-$, where Φ_0 is nonnegative, nondecreasing, and continuous on I. Defining

$$\Gamma_0 = \Psi_0 \circ \Phi_0, \qquad \Psi_0(x) = \int_0^x \psi_0(v) \, dv,$$

and assuming that $\Phi_0(x) < \Phi_0(\mathsf{x}) < u_0$ for every $x < \mathsf{x}$, we have

$$\theta_0(\mathsf{x}) = \partial_- \mathsf{GCM}_{[0, u_0]}(\Gamma_0 \circ \Phi_0^-) \circ \Phi_0(\mathsf{x}) \tag{1}$$

whenever θ_0 is left-continuous at x. In the terminology of Westling and Carone (2020), an estimator of $\theta_0(\mathsf{x})$ is of the Generalized Grenander-type if it is obtained by replacing Γ_0, Φ_0, and u_0 in the preceding display with estimators $\widehat{\Gamma}_n$, $\widehat{\Phi}_n$, and $\widehat{u}_n$ (say); to be specific, an estimator of the Generalized Grenander-type is of the form

$$\widehat{\theta}_n(\mathsf{x}) = \partial_- \mathsf{GCM}_{[0, \widehat{u}_n]}(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^-) \circ \widehat{\Phi}_n(\mathsf{x}),$$

where $\widehat{\Phi}_n$ is nonnegative, nondecreasing, and right-continuous. Of course, Grenander-type estimators are of the generalized Grenander-type (with $\widehat{\Phi}_n$ equal to the identity mapping) whenever the associated estimator of I is of the form $[0, \widehat{u}_n]$, but the class of generalized Grenander-type estimators contains many important estimators that are not of Grenander-type, a canonical example being the celebrated isotonic regression estimator of Brunk (1958).

EXAMPLE 2 (Monotone Regression Estimation). Suppose $(Y_1, X_1), \ldots, (Y_n, X_n)$ are i.i.d. copies of (Y, X), where X is a continuously distributed random variable and where the regression function $\mu_0(x) = \mathbb{E}(Y|X = x)$ is nondecreasing. For x in the interior of the support of X, the Brunk (1958) estimator of $\mu_0(\mathsf{x})$ is

$$\widehat{\mu}_n(\mathsf{x}) = \partial_- \mathsf{GCM}_{[0, 1]}(\widehat{\Gamma}_n \circ \widehat{F}_n^-) \circ \widehat{F}_n(\mathsf{x}),$$

where $\widehat{\Gamma}_n(x) = n^{-1} \sum_{i=1}^{n} Y_i \mathbb{1}(X_i \leq x)$ and where $\widehat{F}_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$ is an estimator of F_0, the cdf of X. Section 5.2 discusses more general monotone regression estimators allowing for covariate adjustment (e.g., Westling, Gilbert and Carone (2020)).

Under regularity conditions, if $\widehat{\theta}_n(\mathsf{x})$ is of the generalized Grenander-type, then its rate of convergence is governed by the flatness of θ_0 around x, as measured by the characteristic exponent

$$\mathfrak{q} = \min\{j \in \mathbb{N} : \partial^j \theta_0(\mathsf{x}) \neq 0\},$$

where $\mathbb{N}$ is the set of positive integers. (When θ_0 is nondecreasing and suitably smooth, $\mathfrak{q}$ is necessarily an odd integer.) To be specific, Westling and Carone (2020, Theorem 3) gave conditions under which

$$(2) \qquad r_n\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) \rightsquigarrow \frac{1}{\partial \Phi_0(\mathsf{x})} \partial_- \mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_\mathsf{x} + \mathcal{M}_\mathsf{x}^{\mathfrak{q}})(0), \qquad r_n = n^{\mathfrak{q}/(1+2\mathfrak{q})},$$

where $\mathcal{G}_\mathsf{x}$ is a scalar multiple of two-sided Brownian motion and where $\mathcal{M}_\mathsf{x}^{\mathfrak{q}}$ is a monomial given by

$$(3) \qquad \mathcal{M}_\mathsf{x}^{\mathfrak{q}}(v) = \frac{\partial^{\mathfrak{q}} \theta_0(\mathsf{x}) \partial \Phi_0(\mathsf{x})}{(\mathfrak{q}+1)!} v^{\mathfrak{q}+1}.$$

In addition to governing the rate of convergence, the characteristic exponent $\mathfrak{q}$ also governs the shape of $\mathcal{M}_\mathsf{x}^{\mathfrak{q}}$. On the other hand, and as the notation suggests, the covariance kernel of $\mathcal{G}_\mathsf{x}$ does not depend on $\mathfrak{q}$. If $\mathfrak{q} = 1$, then the distribution of $\partial_- \mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_\mathsf{x} + \mathcal{M}_\mathsf{x}^{\mathfrak{q}})(0)$ is a scaled Chernoff distribution. More generally, the distribution of $\partial_- \mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_\mathsf{x} + \mathcal{M}_\mathsf{x}^{\mathfrak{q}})(0)$ is a scaled Chernoff-type distribution (in the terminology of Han and Kato (2022)).

Among other things, the following assumption guarantees validity of the representation (1) and ensures existence of the right-hand side of (3).

ASSUMPTION A. For some $\delta > 0$ and some $\mathfrak{s} \geq \mathfrak{q}$, the following are satisfied:

(A1) $I \subseteq \mathbb{R}$ is an interval and $I_\mathsf{x}^{\delta} = \{x \in \mathbb{R} : |x - \mathsf{x}| \leq \delta\} \subseteq I$.

(A2) θ_0 is nondecreasing and bounded on I, and $\lfloor \mathfrak{s} \rfloor$ times continuously differentiable on I_x^{δ} with

$$\sup_{x, x' \in I_\mathsf{x}^{\delta}} \frac{|\partial^{\lfloor \mathfrak{s} \rfloor} \theta_0(x) - \partial^{\lfloor \mathfrak{s} \rfloor} \theta_0(x')|}{|x - x'|^{\mathfrak{s} - \lfloor \mathfrak{s} \rfloor}} < \infty.$$

(A3) Φ_0 is nonnegative, nondecreasing, continuous, and bounded on I, and $\lfloor \mathfrak{s} \rfloor - \mathfrak{q} + 1$ times continuously differentiable on I_x^{δ} with $\partial \Phi_0(\mathsf{x}) \neq 0$ and

$$\sup_{x, x' \in I_\mathsf{x}^{\delta}} \frac{|\partial^{\lfloor \mathfrak{s} \rfloor - \mathfrak{q} + 1} \Phi_0(x) - \partial^{\lfloor \mathfrak{s} \rfloor - \mathfrak{q} + 1} \Phi_0(x')|}{|x - x'|^{\mathfrak{s} - \lfloor \mathfrak{s} \rfloor}} < \infty.$$

3. Bootstrap-assisted distributional approximation.

The distributional approximation (2) motivates a plug-in procedure for statistical inference: with consistent estimators of $\mathcal{M}_\mathsf{x}^{\mathfrak{q}}$ and the covariance kernel of $\mathcal{G}_\mathsf{x}$, the limiting law in (2) can be simulated. This approach requires characterizing the covariance kernel explicitly, and then forming a consistent estimator based on preliminary nonparametric smoothing methods (in addition to the other preliminary nonparametric estimators needed). Also, simulating from the limiting law requires specifying or estimating the unknown characteristic exponent $\mathfrak{q}$. As a consequence, although conceptually feasible, the plug-in approach is often computationally cumbersome to employ in practice.

Since resampling methods have the potential to provide automatic and robust distributional approximations, it seems natural to consider such methods as an alternative to plug-in methods. Letting $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ denote a generic (not necessarily nonparametric) bootstrap analog

of $(\widehat{\Gamma}_n, \widehat{\Phi}_n, \widehat{u}_n)$ and assuming that $\widehat{\Phi}_n^*$ is nondecreasing and right-continuous, the associated bootstrap analog of $\widehat{\theta}_n(\mathsf{x})$ is

$$\widehat{\theta}_n^*(\mathsf{x}) = \partial_-\mathsf{GCM}_{[0,\widehat{u}_n^*]}(\widehat{\Gamma}_n^* \circ \widehat{\Phi}_n^{*-}) \circ \widehat{\Phi}_n^*(\mathsf{x}).$$

As is well documented for the classical Grenander estimator (e.g., Kosorok (2008), Sen, Banerjee and Woodroofe (2010)), the bootstrap analog of (2) does not (necessarily) hold when $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ is obtained by means of the nonparametric bootstrap. In other words,

$$r_n\big(\widehat{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})\big) \not\rightsquigarrow_{\mathbb{P}} \frac{1}{\partial \Phi_0(\mathsf{x})} \partial_-\mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_\mathsf{x} + \mathcal{M}_\mathsf{x}^{\mathsf{q}})(0)$$

in general; that is, the bootstrap is inconsistent in general. (A precise statement is provided in Theorem A.1 in Appendix A.3.) It turns out, however, that under plausible conditions on $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$, a valid bootstrap-based distributional approximation can be obtained by employing

$$\widetilde{\theta}_n^*(\mathsf{x}) = \partial_-\mathsf{GCM}_{[0,\widehat{u}_n^*]}(\widetilde{\Gamma}_n^* \circ \widehat{\Phi}_n^{*-}) \circ \widehat{\Phi}_n^*(\mathsf{x}),$$

where, for some judiciously chosen $\widetilde{M}_{\mathsf{x},n}$,

$$\widetilde{\Gamma}_n^*(x) = \widehat{\Gamma}_n^*(x) - \widehat{\Gamma}_n(x) + \widehat{\theta}_n(\mathsf{x})\widehat{\Phi}_n(x) + \widetilde{M}_{\mathsf{x},n}(x - \mathsf{x}).$$

(As the notation suggests, and as further discussed below, a defining property of $\widetilde{M}_{\mathsf{x},n}$ is that a suitably re-scaled version of it can be interpreted as an estimator of $\mathcal{M}_\mathsf{x}^{\mathsf{q}}$.)

3.1. *Heuristics.* To explain the source of the bootstrap failure and motivate the functional form of $\widetilde{\Gamma}_n^*$, it is useful to begin by sketching the derivation of (2). For any $t \in \mathbb{R}$ and under mild conditions on $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$, it follows from the so-called (generalized) switch relation (e.g., Appendix A.1) that

$$\mathbb{P}\big[r_n\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) \le t\big] = \mathbb{P}\Big[\underset{v \in \widehat{V}_{\mathsf{x},n}^{\mathsf{q}}}{\operatorname{argmin}}\{\widehat{G}_{\mathsf{x},n}^{\mathsf{q}}(v) + M_{\mathsf{x},n}^{\mathsf{q}}(v) - t\widehat{L}_{\mathsf{x},n}^{\mathsf{q}}(v)\} \ge \widehat{Z}_{\mathsf{x},n}^{\mathsf{q}}\Big],$$

where $\widehat{V}_{\mathsf{x},n}^{\mathsf{q}}$ is a "large" subset of $\mathbb{R}$, $\widehat{G}_{\mathsf{x},n}^{\mathsf{q}}$ is a stochastic process given by

$$\widehat{G}_{\mathsf{x},n}^{\mathsf{q}}(v) = \sqrt{na_n}\big[\widehat{\Gamma}_n\big(\mathsf{x} + va_n^{-1}\big) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0\big(\mathsf{x} + va_n^{-1}\big) + \Gamma_0(\mathsf{x})\big]$$

$$- \theta_0(\mathsf{x})\sqrt{na_n}\big[\widehat{\Phi}_n\big(\mathsf{x} + va_n^{-1}\big) - \widehat{\Phi}_n(\mathsf{x}) - \Phi_0\big(\mathsf{x} + va_n^{-1}\big) + \Phi_0(\mathsf{x})\big],$$

$$a_n = n^{1/(1+2\mathsf{q})},$$

$M_{\mathsf{x},n}^{\mathsf{q}}$ is a smooth nonrandom function, $\widehat{L}_{\mathsf{x},n}^{\mathsf{q}}$ can be interpreted as an estimator of the linear function $\mathcal{L}_\mathsf{x}$ given by $\mathcal{L}_\mathsf{x}(v) = v\partial\Phi_0(\mathsf{x})$, and where $\widehat{Z}_{\mathsf{x},n}^{\mathsf{q}}$ is zero when $\widehat{\Phi}_n$ is the identity mapping and can be interpreted as an "estimator" of zero more generally. (The exact definitions of $\widehat{V}_{\mathsf{x},n}^{\mathsf{q}}$, $M_{\mathsf{x},n}^{\mathsf{q}}$, $\widehat{L}_{\mathsf{x},n}^{\mathsf{q}}$, and $\widehat{Z}_{\mathsf{x},n}^{\mathsf{q}}$ are given in Appendix A.4.) In the above and the sequel, the superscript q indicates that the superscripted object is defined on a "localized" domain.

Under mild conditions, $\widehat{G}_{\mathsf{x},n}^{\mathsf{q}}$ converges weakly to a scalar multiple of two-sided Brownian motion $\mathcal{G}_\mathsf{x}$ (say), while $M_{\mathsf{x},n}^{\mathsf{q}}$, $\widehat{L}_{\mathsf{x},n}^{\mathsf{q}}$, and $\widehat{Z}_{\mathsf{x},n}^{\mathsf{q}}$ converge in probability to $\mathcal{M}_\mathsf{x}^{\mathsf{q}}$, $\mathcal{L}_\mathsf{x}$, and zero, respectively. In other words, we would expect that

$$(4) \qquad\qquad (\widehat{G}_{\mathsf{x},n}^{\mathsf{q}}, \widehat{L}_{\mathsf{x},n}^{\mathsf{q}}, \widehat{Z}_{\mathsf{x},n}^{\mathsf{q}}) \rightsquigarrow (\mathcal{G}_\mathsf{x}, \mathcal{L}_\mathsf{x}, 0),$$

and that

$$(5) \qquad\qquad M_{\mathsf{x},n}^{\mathsf{q}} \rightsquigarrow \mathcal{M}_\mathsf{x}^{\mathsf{q}}.$$

It therefore stands to reason that under mild additional conditions (including suitable convergence of $\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}$ to $\mathbb{R}$), we have

$$\lim_{n\to\infty}\mathbb{P}\big[r_n\big(\widehat{\theta}_n(\mathsf{x})-\theta_0(\mathsf{x})\big)\le t\big]=\mathbb{P}\Big[\operatorname*{argmin}_{v\in\mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v)+\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}(v)-t\mathcal{L}_{\mathsf{x}}(v)\}\ge 0\Big]$$

$$=\mathbb{P}\Big[\frac{1}{\partial\Phi_0(\mathsf{x})}\partial_-\mathsf{GCM}_{\mathbb{R}}\big(\mathcal{G}_{\mathsf{x}}+\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}\big)(0)\le t\Big],$$

where the first equality follows from an argmax continuous mapping theorem, and the second equality is obtained by another application of the switch relation.

Next, consider the bootstrap analog of $\widehat{\theta}_n(\mathsf{x})$. For any $t\in\mathbb{R}$, the switch relation gives

$$\mathbb{P}^*_n\big[r_n\big(\widehat{\theta}^*_n(\mathsf{x})-\widehat{\theta}_n(\mathsf{x})\big)\le t\big]=\mathbb{P}^*_n\Big[\operatorname*{argmin}_{v\in\widehat{V}^{\mathfrak{q},*}_{\mathsf{x},n}}\{\widehat{G}^{\mathfrak{q},*}_{\mathsf{x},n}(v)+\widehat{M}^{\mathfrak{q}}_{\mathsf{x},n}(v)-t\widehat{L}^{\mathfrak{q},*}_{\mathsf{x},n}(v)\}\ge\widehat{Z}^{\mathfrak{q},*}_{\mathsf{x},n}\Big],$$

where

$$\widehat{G}^{\mathfrak{q},*}_{\mathsf{x},n}(v)=\sqrt{na_n}\big[\widehat{\Gamma}^*_n\big(\mathsf{x}+va_n^{-1}\big)-\widehat{\Gamma}^*_n(\mathsf{x})-\widehat{\Gamma}_n\big(\mathsf{x}+va_n^{-1}\big)+\widehat{\Gamma}_n(\mathsf{x})\big]$$

$$-\widehat{\theta}_n(\mathsf{x})\sqrt{na_n}\big[\widehat{\Phi}^*_n\big(\mathsf{x}+va_n^{-1}\big)-\widehat{\Phi}^*_n(\mathsf{x})-\widehat{\Phi}_n\big(\mathsf{x}+va_n^{-1}\big)+\widehat{\Phi}_n(\mathsf{x})\big],$$

and where $\widehat{V}^{\mathfrak{q},*}_{\mathsf{x},n}$, $\widehat{M}^{\mathfrak{q}}_{\mathsf{x},n}$, $\widehat{L}^{\mathfrak{q},*}_{\mathsf{x},n}$, and $\widehat{Z}^{\mathfrak{q},*}_{\mathsf{x},n}$ are bootstrap analogs of $\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}$, $M^{\mathfrak{q}}_{\mathsf{x},n}$, $\widehat{L}^{\mathfrak{q}}_{\mathsf{x},n}$, and $\widehat{Z}^{\mathfrak{q}}_{\mathsf{x},n}$, respectively. (The exact definitions of $\widehat{V}^{\mathfrak{q},*}_{\mathsf{x},n}$, $\widehat{M}^{\mathfrak{q}}_{\mathsf{x},n}$, $\widehat{L}^{\mathfrak{q},*}_{\mathsf{x},n}$, and $\widehat{Z}^{\mathfrak{q},*}_{\mathsf{x},n}$ are given in Appendix A.4.) If (4) holds, then the following bootstrap counterpart thereof can also be expected to hold:

$$(6)\qquad\qquad\big(\widehat{G}^{\mathfrak{q},*}_{\mathsf{x},n},\widehat{L}^{\mathfrak{q},*}_{\mathsf{x},n},\widehat{Z}^{\mathfrak{q},*}_{\mathsf{x},n}\big)\rightsquigarrow_{\mathbb{P}}(\mathcal{G}_{\mathsf{x}},\mathcal{L}_{\mathsf{x}},0).$$

On the other hand, unlike $M^{\mathfrak{q}}_{\mathsf{x},n}$ the process $\widehat{M}^{\mathfrak{q}}_{\mathsf{x},n}$ is (random and) nonsmooth, so the bootstrap counterpart of (5) typically fails, implying in turn that the bootstrap is inconsistent; for details, see Theorem A.1 (stated in Appendix A.3) and its proof.

In other words, the sole source of the bootstrap inconsistency is the failure of $\widehat{M}^{\mathfrak{q}}_{\mathsf{x},n}$ to replicate the salient properties of $M^{\mathfrak{q}}_{\mathsf{x},n}$. Recognizing this, our proposed estimator $\widetilde{\theta}^*_n(\mathsf{x})$ has been carefully constructed so that it differs from $\widehat{\theta}^*_n(\mathsf{x})$ only in terms of the implied estimator of $\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$. To be specific, $\widetilde{\theta}^*_n(\mathsf{x})$ is similar to $\widehat{\theta}^*_n(\mathsf{x})$ insofar as it satisfies the following switch relation: For any $t\in\mathbb{R}$,

$$\mathbb{P}^*_n\big[r_n\big(\widetilde{\theta}^*_n(\mathsf{x})-\widehat{\theta}_n(\mathsf{x})\big)\le t\big]=\mathbb{P}^*_n\Big[\operatorname*{argmin}_{v\in\widehat{V}^{\mathfrak{q},*}_{\mathsf{x},n}}\{\widehat{G}^{\mathfrak{q},*}_{\mathsf{x},n}(v)+\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}(v)-t\widehat{L}^{\mathfrak{q},*}_{\mathsf{x},n}(v)\}\ge\widehat{Z}^{\mathfrak{q},*}_{\mathsf{x},n}\Big],$$

where $\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}$ is the following transformation of $\widetilde{M}_{\mathsf{x},n}$:

$$\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}(v)=\sqrt{na_n}\widetilde{M}_{\mathsf{x},n}\big(va_n^{-1}\big).$$

As a consequence, if (6) holds and if $\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}\rightsquigarrow_{\mathbb{P}}\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$, then it stands to reason that under mild additional conditions $\widetilde{\theta}^*_n(\mathsf{x})$ satisfies the following bootstrap counterpart of (2):

$$(7)\qquad\qquad r_n\big(\widetilde{\theta}^*_n(\mathsf{x})-\widehat{\theta}_n(\mathsf{x})\big)\rightsquigarrow_{\mathbb{P}}\frac{1}{\partial\Phi_0(\mathsf{x})}\partial_-\mathsf{GCM}_{\mathbb{R}}\big(\mathcal{G}_{\mathsf{x}}+\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}\big)(0).$$

In words, our proposed bootstrap-assisted inference procedure can be explained as follows: we employ the bootstrap to approximate the "fluctuations" of the process $\mathcal{G}_{\mathsf{x}}$, while we use the estimator $\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}$ to approximate $\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$. It is an interesting question whether our approach offers provably better distributional approximations than the plug-in procedure mentioned previously, but it is beyond the scope of this paper to answer that question.

3.2. *Main result.* Our heuristic derivation of (2) can be made rigorous by providing conditions under which four properties hold. First, the switch relation(s) must be valid. Second, the convergence properties (4) and (5) must hold. Third, to use (4) and (5) to obtain the result

$$\operatorname*{argmin}_{v \in \widehat{V}^{\mathsf{q}}_{\mathsf{x},n}}\{\widehat{G}^{\mathsf{q}}_{\mathsf{x},n}(v) + M^{\mathsf{q}}_{\mathsf{x},n}(v) - t\widehat{L}^{\mathsf{q}}_{\mathsf{x},n}(v)\} \rightsquigarrow \operatorname*{argmin}_{v \in \mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}^{\mathsf{q}}_{\mathsf{x}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\}$$

with the help of a suitable argmax continuous mapping theorem, tightness of the left-hand side in the previous display must hold. Finally, to furthermore obtain the conclusion

$$\lim_{n \to \infty} \mathbb{P}\Big[\operatorname*{argmin}_{v \in \widehat{V}^{\mathsf{q}}_{\mathsf{x},n}}\{\widehat{G}^{\mathsf{q}}_{\mathsf{x},n}(v) + M^{\mathsf{q}}_{\mathsf{x},n}(v) - t\widehat{L}^{\mathsf{q}}_{\mathsf{x},n}(v)\} \geq \widehat{Z}^{\mathsf{q}}_{\mathsf{x},n}\Big]$$

$$= \mathbb{P}\Big[\operatorname*{argmin}_{v \in \mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}^{\mathsf{q}}_{\mathsf{x}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\} \geq 0\Big],$$

the cdf of $\operatorname{argmin}_{v \in \mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}^{\mathsf{q}}_{\mathsf{x}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\}$ must be continuous at zero.

Conditions under which the second and third properties hold can be formulated with the help of well-known empirical process results. We base our formulations on van der Vaart and Wellner (1996) and employ the generalized argmax continuous mapping theorem of Cox (2022), which applies also to settings where the feasibility set varies with n. The first and fourth properties, on the other hand, seem more difficult to verify. Regarding the first property, it turns out that the generalization of the switch relation employed by Westling and Carone (2020) requires additional conditions in order to be valid. To address this concern, we present a modified generalized switch relation whose assumptions include a condition not present in Lemma 1 of Westling and Carone (2020, Supplement). Thankfully, the condition in question seems very mild and having imposed it we are able to preserve the main implication of Lemma 1 of Westling and Carone (2020, Supplement); for details, see Lemma A.1 in Appendix A.1. In the special case where $\mathsf{q} = 1$, the fourth property follows from well known properties of the Chernoff distribution. More generally, however, we are unaware of existing results guaranteeing the requisite continuity property when $\mathsf{q} \neq 1$, but fortunately it turns out that the continuity property of interest can be established; for details, see Lemma A.2 in Appendix A.2.

The following assumption collects the conditions under which our verification of the four above-mentioned properties will proceed.

ASSUMPTION B. For some $\delta > 0$, the following are satisfied:

(B1) $\widehat{G}^{\mathsf{q}}_{\mathsf{x},n} \rightsquigarrow \mathcal{G}_{\mathsf{x}}$ and $\widehat{G}^{\mathsf{q},*}_{\mathsf{x},n} \rightsquigarrow_{\mathbb{P}} \mathcal{G}_{\mathsf{x}}$, where $\mathcal{G}_{\mathsf{x}}$ is a centered Gaussian process whose covariance function is $\mathcal{C}_{\mathsf{x}}(s, t) = C(|s| \wedge |t|)\mathbb{1}(\operatorname{sign}(s) = \operatorname{sign}(t))$ for some $C > 0$.

(B2) There exist $\beta < \mathsf{q} + 1$ and events $\mathcal{A}_n$ with $\lim_{n \to \infty} \mathbb{P}[\mathcal{A}_n] = 1$,

$$\sup_{V \in [1, a_n \delta]} \mathbb{E}\Big[V^{-\beta} \sup_{|v| \in [V, 2V]} |\widehat{G}^{\mathsf{q}}_{\mathsf{x},n}(v)|\mathbb{1}_{\mathcal{A}_n}\Big] = O(1)$$

and

$$\sup_{V \in [1, a_n \delta]} \mathbb{E}\Big[V^{-\beta} \sup_{|v| \in [V, 2V]} |\widehat{G}^{\mathsf{q},*}_{\mathsf{x},n}(v)|\mathbb{1}_{\mathcal{A}_n}\Big] = O(1).$$

(B3) $\sup_{x \in I} |\widehat{\Gamma}_n(x) - \Gamma_0(x)| = o_{\mathbb{P}}(1)$ and $\sup_{x \in I} |\widehat{\Gamma}^*_n(x) - \widehat{\Gamma}_n(x)| = o_{\mathbb{P}}(1)$.

(B4) $\sup_{x \in I} |\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1)$ and $\sup_{x \in I} |\widehat{\Phi}^*_n(x) - \widehat{\Phi}_n(x)| = o_{\mathbb{P}}(1)$. In addition,

$$a_n \sup_{x \in I^\delta_{\mathsf{x}}} |\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1) \quad \text{and} \quad a_n \sup_{x \in I^\delta_{\mathsf{x}}} |\widehat{\Phi}^*_n(x) - \widehat{\Phi}_n(x)| = o_{\mathbb{P}}(1).$$

(B5) For some $u_0 > \Phi_0(\mathsf{x})$, $\widehat{u}_n \geq u_0 + o_{\mathbb{P}}(1)$ and $\widehat{u}^*_n \geq \widehat{u}_n + o_{\mathbb{P}}(1)$.

(B6) $\widehat{\Phi}_n$ and $\widehat{\Phi}_n^*$ are nonnegative, nondecreasing, and right-continuous on I. In addition, $\{0, \widehat{u}_n\} \subseteq \widehat{\Phi}_n(I)$ and $\{0, \widehat{u}_n^*\} \subseteq \widehat{\Phi}_n^*(I)$. Also, $\widehat{\Phi}_n(I) \cap [0, \widehat{u}_n]$ and $\widehat{\Phi}_n^*(I) \cap [0, \widehat{u}_n^*]$ are closed.

(B7)

$$\sqrt{na_n} \sup_{x \in I_x^\delta} \left| \widehat{\Phi}_n(x) - \widehat{\Phi}_n(x-) \right| = o_\mathbb{P}(1) \quad \text{and} \quad \sqrt{na_n} \sup_{x \in I_x^\delta} \left| \widehat{\Phi}_n^*(x) - \widehat{\Phi}_n^*(x-) \right| = o_\mathbb{P}(1).$$

Verification of the bootstrap parts of Assumption B will be discussed in Section 4.2 below. When combined with Assumption A, Assumption B suffices in order to establish (2) and (6). In addition, (7) can be shown to hold if $\widetilde{M}_{x,n}$ satisfies the following.

ASSUMPTION C. $\widetilde{M}_{x,n}^q \rightsquigarrow_\mathbb{P} \mathcal{M}_x^q$ and, for some $c > 0$ and every $K > 0$,

$$\liminf_{n \to \infty} \mathbb{P}\left[\inf_{|v| > K^{-1}} \widetilde{M}_{x,n}(v) \geq cK^{-(q+1)} \right] = 1.$$

Moreover, continuity of the cdf of $\partial_- \mathsf{GCM}_\mathbb{R}(\mathcal{G}_x + \mathcal{M}_x^q)(0)$ can be shown with the help of the arguments used to show that the cdf of $\operatorname{argmin}_{v \in \mathbb{R}}\{\mathcal{G}_x(v) + \mathcal{M}_x^q(v) - t\mathcal{L}_x(v)\}$ is continuous at zero. As a consequence, we obtain the following result.

THEOREM 1. *Suppose Assumptions A, B, and C are satisfied. Then* (2) *and* (7) *hold, and*

$$(8) \qquad \sup_{t \in \mathbb{R}} \left| \mathbb{P}_n^*\left[\widetilde{\theta}_n^*(x) - \widehat{\theta}_n(x) \leq t \right] - \mathbb{P}\left[\widehat{\theta}_n(x) - \theta_0(x) \leq t \right] \right| = o_\mathbb{P}(1).$$

In an attempt to emphasize the rate-adaptive nature of the consistency property enjoyed by the bootstrap-based distributional approximation based on $\widetilde{\theta}_n^*(x)$, the formulation (8) deliberately omits the rate term r_n present in (2) and (7). Theorem 1 has immediate implications for inference. For instance, it follows from (2) and (8) that for any $\alpha \in (0, 1)$, we have

$$\lim_{n \to \infty} \mathbb{P}\left[\theta_0(x) \in \mathsf{CI}_{1-\alpha,n}(x) \right] = 1 - \alpha,$$

where, defining $Q_{a,n}^*(x) = \inf\{Q \in \mathbb{R} : \mathbb{P}_n^*[\widetilde{\theta}_n^*(x) - \widehat{\theta}_n(x) \leq Q] \geq a\}$,

$$\mathsf{CI}_{1-\alpha,n}(x) = \left[\widehat{\theta}_n(x) - Q_{1-\alpha/2,n}^*(x), \widehat{\theta}_n(x) - Q_{\alpha/2,n}^*(x) \right]$$

is the (nominal) level $1 - \alpha$ bootstrap confidence interval for θ_0 based on the "percentile method" (in the terminology of van der Vaart (1998)).

4. Implementation. Suppose Assumption A is satisfied and suppose $(\widehat{\Gamma}_n, \widehat{\Phi}_n, \widehat{u}_n)$ satisfies the nonbootstrap parts of Assumption B. Then, in order to compute the estimator $\widetilde{\theta}_n^*(x)$ upon which our proposed bootstrap-assisted distributional approximation is based, two implementation issues must be addressed, namely the choice/construction of $\widetilde{M}_{x,n}$ and $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$, respectively. Section 4.1 demonstrates the plausibility of Assumption C by exhibiting estimators $\widetilde{M}_{x,n}$ satisfying it under Assumptions A and D, the latter being a high-level condition that typically holds whenever the nonbootstrap parts of Assumption B hold. Then, Section 4.2 exhibits easy-to-compute $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ satisfying the bootstrap parts of Assumption B under a random sampling assumption and other mild conditions.

4.1. *Mean function estimation.* The ease with which an $\widetilde{M}_{\mathsf{x},n}$ satisfying Assumption C can be constructed depends in particular on whether q is known. To facilitate the subsequent discussion, let $\mathcal{D}_0(\mathsf{x}) = 0$ and for $j = 1, \ldots, \lfloor \mathsf{s} \rfloor$, define (recursively)

$$\mathcal{D}_j(\mathsf{x}) = \lim_{x \to \mathsf{x}} \frac{\Upsilon_0(x) - \Upsilon_0(\mathsf{x}) - \sum_{k=0}^{j-1}[(k+1)!]^{-1}\mathcal{D}_k(\mathsf{x})(x - \mathsf{x})^{k+1}}{(x - \mathsf{x})^{j+1}},$$

$$\Upsilon_0 = \Gamma_0 - \theta_0(\mathsf{x})\Phi_0.$$

Noting that $\partial\Gamma_0(x) = \theta_0(x)\partial\Phi_0(x)$ for x near x, we have $\mathcal{D}_0(\mathsf{x}) = \cdots = \mathcal{D}_{\mathsf{q}-1}(\mathsf{x}) = 0$ and

$$\mathcal{D}_j(\mathsf{x}) = \frac{1}{(j+1)!} \sum_{k=\mathsf{q}}^{j} \binom{j}{k} \partial^k \theta_0(\mathsf{x}) \partial^{j+1-k} \Phi_0(\mathsf{x}), \quad j = \mathsf{q}, \ldots, \lfloor \mathsf{s} \rfloor.$$

In particular, $\mathcal{D}_{\mathsf{q}}(\mathsf{x}) = \partial^{\mathsf{q}}\theta_0(\mathsf{x})\partial\Phi_0(\mathsf{x})/(\mathsf{q}+1)!$ and therefore $\mathcal{M}_{\mathsf{x}}^{\mathsf{q}}(v) = \mathcal{D}_{\mathsf{q}}(\mathsf{x})v^{\mathsf{q}+1}$.

4.1.1. *Mean function estimation with known* q. First, consider the (simpler) case where q is known. In this case, if

(9) $$\widetilde{\mathcal{D}}_{\mathsf{q},n}(\mathsf{x}) \to_{\mathbb{P}} \mathcal{D}_{\mathsf{q}}(\mathsf{x}),$$

then Assumption C holds when

(10) $$\widetilde{M}_{\mathsf{x},n}(x) = \widetilde{\mathcal{D}}_{\mathsf{q},n}(\mathsf{x})x^{\mathsf{q}+1}.$$

Example-specific estimators $\widetilde{\mathcal{D}}_{\mathsf{q},n}(\mathsf{x})$ satisfying the consistency requirement (9) may be readily available. For instance, in the case of the Grenander (1956) estimator described in Example 1, we have $\mathcal{D}_{\mathsf{q}}(\mathsf{x}) = \partial^{\mathsf{q}} f_0(\mathsf{x})/(\mathsf{q}+1)!$, a consistent estimator of which can be based on any consistent estimator of $\partial^{\mathsf{q}} f_0(\mathsf{x})$, such as a standard kernel estimator or, if the evaluation point x is near the boundary of the support of X, boundary adaptive versions thereof.

More generic estimators are also available. For specificity, the remainder of this section focuses on estimators of $\mathcal{D}_j(\mathsf{x})$ obtained by applying numerical derivative-type operators to the following (possibly) nonsmooth estimator of Υ_0:

$$\widehat{\Upsilon}_n = \widehat{\Gamma}_n - \widehat{\theta}_n(\mathsf{x})\widehat{\Phi}_n.$$

The fact that $\mathcal{D}_j(\mathsf{x}) = 0$ for $j < \mathsf{q}$ implies that $\mathcal{D}_{\mathsf{q}}(\mathsf{x})$ admits the "monomial approximation" representation

$$\mathcal{D}_{\mathsf{q}}(\mathsf{x}) = \lim_{\epsilon \to 0} \{\epsilon^{-(\mathsf{q}+1)}[\Upsilon_0(\mathsf{x}+\epsilon) - \Upsilon_0(\mathsf{x})]\},$$

motivating the estimator

$$\widetilde{\mathcal{D}}_{\mathsf{q},n}^{\mathrm{MA}}(\mathsf{x}) = \epsilon_n^{-(\mathsf{q}+1)}[\widehat{\Upsilon}_n(\mathsf{x}+\epsilon_n) - \widehat{\Upsilon}_n(\mathsf{x})],$$

where $\epsilon_n > 0$ is a tuning parameter. Similarly, the generic "forward difference" representation

$$\mathcal{D}_{\mathsf{q}}(\mathsf{x}) = \lim_{\epsilon \to 0} \left\{ \epsilon^{-(\mathsf{q}+1)} \sum_{k=1}^{\mathsf{q}+1} (-1)^{k+\mathsf{q}+1} \binom{\mathsf{q}+1}{k} [\Upsilon_0(\mathsf{x}+k\epsilon) - \Upsilon_0(\mathsf{x})] \right\}$$

motivates the estimator

$$\widetilde{\mathcal{D}}_{\mathsf{q},n}^{\mathrm{FD}}(\mathsf{x}) = \epsilon_n^{-(\mathsf{q}+1)} \sum_{k=1}^{\mathsf{q}+1} (-1)^{k+\mathsf{q}+1} \binom{\mathsf{q}+1}{k} [\widehat{\Upsilon}_n(\mathsf{x}+k\epsilon_n) - \widehat{\Upsilon}_n(\mathsf{x})].$$

For $\eta > 0$, let

$$\widehat{G}_{\mathsf{x},n}(v; \eta) = \sqrt{n\eta^{-1}}\big[\widehat{\Gamma}_n(\mathsf{x} + v\eta) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x} + v\eta) + \Gamma_0(\mathsf{x})\big]$$
$$- \theta_0(\mathsf{x})\sqrt{n\eta^{-1}}\big[\widehat{\Phi}_n(\mathsf{x} + v\eta) - \widehat{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x} + v\eta) + \Phi_0(\mathsf{x})\big],$$

and

$$\widehat{R}_{\mathsf{x},n}(v; \eta) = \eta^{-1}\big[\widehat{\Phi}_n(\mathsf{x} + v\eta) - \widehat{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x} + v\eta) + \Phi_0(\mathsf{x})\big].$$

Using this notation, the first part of (B1) can be restated as

$$\widehat{G}_{\mathsf{x},n}(\cdot; a_n^{-1}) \rightsquigarrow \mathcal{G}_{\mathsf{x}}$$

and the displayed condition of (B4) implies that

$$\widehat{R}_{\mathsf{x},n}(\cdot; a_n^{-1}) = o_{\mathbb{P}}(1).$$

In the preceding displays, one can typically replace $a_n^{-1} = n^{-1/(1+2\mathfrak{q})}$ by any $\eta_n > 0$ with $\eta_n = o(1)$ and $a_n^{-1}\eta_n^{-1} = O(1)$. As a consequence, validity of the following assumption usually follows as a by-product of the arguments used to justify (B1) and (B4).

ASSUMPTION D. For every $\eta_n > 0$ with $\eta_n = o(1)$ and $a_n^{-1}\eta_n^{-1} = O(1)$,

$$\widehat{G}_{\mathsf{x},n}(1; \eta_n) = O_{\mathbb{P}}(1) \quad \text{and} \quad \widehat{R}_{\mathsf{x},n}(1; \eta_n) = O_{\mathbb{P}}(1).$$

In turn, Assumption D is useful for the purposes of analyzing $\widetilde{\mathcal{D}}_{j,n}^{\mathrm{MA}}$, and $\widetilde{\mathcal{D}}_{j,n}^{\mathrm{FD}}$.

LEMMA 1. *Suppose Assumptions* A *and* D *are satisfied, that* $r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})) = O_{\mathbb{P}}(1)$, *and that* $\epsilon_n \to 0$ *and* $n\epsilon_n^{1+2\mathfrak{q}} \to \infty$. *If* $\widetilde{\mathcal{D}}_{\mathfrak{q},n} \in \{\widetilde{\mathcal{D}}_{\mathfrak{q},n}^{\mathrm{MA}}, \widetilde{\mathcal{D}}_{\mathfrak{q},n}^{\mathrm{FD}}\}$, *then* (9) *is satisfied and Assumption* C *holds with* $\widetilde{M}_{\mathsf{x},n}$ *in* (10).

4.1.2. *Mean function estimation*: *Adaptation to unknown degeneracy.* Next, consider the somewhat more complicated case where $\mathfrak{q}$ is unknown. In this case, we are able to obtain a positive result under the (additional) assumption that there are known integers $\bar{\mathfrak{q}}$ and $\underline{\mathfrak{s}}$ for which $\mathfrak{q} \le \bar{\mathfrak{q}} \le \underline{\mathfrak{s}} < \mathfrak{s}$. As the notation suggests, $\bar{\mathfrak{q}}$ and $\underline{\mathfrak{s}}$ are upper and lower bounds on $\mathfrak{q}$ and $\mathfrak{s}$, respectively. The stated condition furthermore implies that $\mathfrak{s} > \mathfrak{q}$, a stronger smoothness condition than the condition $\mathfrak{s} \ge \mathfrak{q}$ of Assumption A.

For any $\bar{\mathfrak{q}} \ge \mathfrak{q}$, noting that $\mathcal{D}_j(\mathsf{x}) = 0$ for $j < \mathfrak{q}$ and that $\mathcal{D}_{\mathfrak{q}}(\mathsf{x}) > 0$ with $\mathfrak{q}$ an odd integer, $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$ can be written as

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \sum_{\ell=1}^{\lfloor(\bar{\mathfrak{q}}+1)/2\rfloor} \mathbb{1}(2\ell - 1 \le \mathfrak{q})\mathcal{D}_{2\ell-1}(\mathsf{x})^+ v^{2\ell}.$$

Dropping the indicator function term from each summand, we obtain

$$\bar{\mathcal{M}}_{\mathsf{x}}(v) = \sum_{\ell=1}^{\lfloor(\bar{\mathfrak{q}}+1)/2\rfloor} \mathcal{D}_{2\ell-1}(\mathsf{x})^+ v^{2\ell}.$$

The majorant $\bar{\mathcal{M}}_{\mathsf{x}}$ is an "adaptive" approximation to $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$ in the sense that it does not depend on $\mathfrak{q}$, yet satisfies the local approximation property

$$\bar{\mathcal{M}}_{\mathsf{x}}^{\mathfrak{q}} \rightsquigarrow \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}, \qquad \bar{\mathcal{M}}_{\mathsf{x}}^{\mathfrak{q}}(v) = \sqrt{na_n}\bar{\mathcal{M}}_{\mathsf{x}}(va_n^{-1}).$$

Moreover, the following ("global") positivity property automatically holds: for some $c > 0$:

$$\inf_{|v| > K^{-1}} \bar{\mathcal{M}}_{\mathsf{x}}(v) > cK^{-(\mathfrak{q}+1)} \quad \text{for every } K > 0.$$

As a consequence, it seems plausible that a "plug-in" estimator of $\bar{\mathcal{M}}_x$ would satisfy Assumption C under reasonable conditions. Indeed, if

$$(11) \qquad a_n^{q-(2\ell-1)}\big(\widetilde{\mathcal{D}}_{2\ell-1,n}(x) - \mathcal{D}_{2\ell-1}(x)\big) = o_{\mathbb{P}}(1), \quad \ell = 1, \dots, \lfloor(\bar{q}+1)/2\rfloor,$$

then Assumption C is satisfied by

$$(12) \qquad \widetilde{M}_{x,n}(v) = \sum_{\ell=1}^{\lfloor(\bar{q}+1)/2\rfloor} \widetilde{\mathcal{D}}_{2\ell-1,n}(x)^+ v^{2\ell}.$$

For "small" ℓ (namely, for $2\ell - 1 < q$), the precision requirement (11) is stronger than consistency, but fortunately it turns out that the requirement can be met as long as $\mathfrak{s}$ admits an integer $\underline{\mathfrak{s}}$ satisfying the condition $\bar{q} \leq \underline{\mathfrak{s}} < \mathfrak{s}$. To describe estimators satisfying (11) under this condition, let the constants $c_1, \dots, c_{\underline{\mathfrak{s}}+1}$ be such that invertibility holds on the part of the $(\underline{\mathfrak{s}}+1) \times (\underline{\mathfrak{s}}+1)$ matrix with element (k, p) given by c_k^p and let the defining property of $\{\lambda_j^{\mathrm{BR}}(k) : 1 \leq k \leq \underline{\mathfrak{s}}+1\}$ be

$$\sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathrm{BR}}(k)c_k^p = \mathbb{1}(p = j+1), \quad p = 1, \dots, \underline{\mathfrak{s}}+1.$$

Then, for any $j = 1, \dots, \underline{\mathfrak{s}}$, the "bias-reduced" estimator

$$\widetilde{\mathcal{D}}_{j,n}^{\mathrm{BR}}(x) = \epsilon_n^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathrm{BR}}(k)\big[\widehat{\Upsilon}_n(x + c_k\epsilon_n) - \widehat{\Upsilon}_n(x)\big]$$

is motivated by the fact that as $\epsilon \to 0$,

$$\mathcal{D}_j(x) = \epsilon^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathrm{BR}}(k)\big[\Upsilon_0(x + c_k\epsilon) - \Upsilon_0(x)\big] + O\big(\epsilon^{\min(\underline{\mathfrak{s}}+1,\mathfrak{s})-j}\big)$$

when $\mathfrak{s} > j$. Relative to $\widetilde{\mathcal{D}}_{j,n}^{\mathrm{MA}}(x)$ and $\widetilde{\mathcal{D}}_{j,n}^{\mathrm{FD}}(x)$, this is a distinguishing feature of $\widetilde{\mathcal{D}}_{j,n}^{\mathrm{BR}}(x)$ and as it turns out this feature will enable us to formulate sufficient conditions for (11).

LEMMA 2. *Suppose Assumptions A and D are satisfied, that $r_n(\widehat{\theta}_n(x) - \theta_0(x)) = O_{\mathbb{P}}(1)$, and that $\epsilon_n \to 0$. If $q \leq \bar{q} \leq \underline{\mathfrak{s}} < \mathfrak{s}$, then*

$$a_n^{q-j}\big(\widetilde{\mathcal{D}}_{j,n}^{\mathrm{BR}}(x) - \mathcal{D}_j(x)\big) = O\big(a_n^{q-j}\epsilon_n^{\min(\underline{\mathfrak{s}}+1,\mathfrak{s})-j}\big) + O_{\mathbb{P}}\big[(a_n\epsilon_n)^{-(j+1/2)} + (a_n\epsilon_n)^{-j}\big]$$

for $j = 1, \dots, \underline{\mathfrak{s}}$, implying in particular that if also

$$n\epsilon_n^{(1+2\bar{q})\min(\underline{\mathfrak{s}},\mathfrak{s}-1)/(\bar{q}-1)} \to 0 \quad \text{and} \quad n\epsilon_n^{1+2\bar{q}} \to \infty,$$

then (11) is satisfied and Assumption C holds with $\widetilde{M}_{x,n}$ in (12) if $\widetilde{\mathcal{D}}_{2\ell-1,n} = \widetilde{\mathcal{D}}_{2\ell-1,n}^{\mathrm{BR}}$.

As alluded to previously, the ability of $\widetilde{\mathcal{D}}^{\mathrm{BR}}$ to satisfy (11) is attributable in large part to its bias properties. In an attempt to highlight this, the first display of the lemma gives a stochastic expansion wherein the $O(a_n^{q-j}\epsilon_n^{\min(\underline{\mathfrak{s}}+1,\mathfrak{s})-j})$ term is a (possibly) negligible bias term. For $\widetilde{\mathcal{D}} \in \{\widetilde{\mathcal{D}}^{\mathrm{MA}}, \widetilde{\mathcal{D}}^{\mathrm{FD}}\}$, the analogous stochastic expansions are of the form

$$a_n^{q-j}\big(\widetilde{\mathcal{D}}_{j,n}(x) - \mathcal{D}_j(x)\big) = O\big(a_n^{q-j}\epsilon_n^{q-j}\big) + O_{\mathbb{P}}\big[(a_n\epsilon_n)^{-(j+1/2)} + (a_n\epsilon_n)^{-j}\big],$$

the $O(a_n^{q-j}\epsilon_n^{q-j})$ term also being a bias term. When $a_n\epsilon_n \to \infty$ (as is required for the "noise" term in the stochastic expansion to be $o_{\mathbb{P}}(1)$), this bias term is nonnegligible and the estimators $\widetilde{\mathcal{D}}^{\mathrm{MA}}$ and $\widetilde{\mathcal{D}}^{\mathrm{FD}}$ therefore do not satisfy (11).

Under additional assumptions (including $\mathfrak{s} \geq \underline{\mathfrak{s}} + 1$ and additional smoothness on Φ_0), $\widetilde{\mathcal{D}}^{\mathrm{BR}}$ admits a Nagar-type mean squared error (MSE) expansion that can be used to select ϵ_n. The resulting approximate MSE formula is of the form

$$\epsilon_n^{2(\underline{\mathfrak{s}}+1-j)} \mathsf{B}_j^{\mathrm{BR}}(\mathsf{x})^2 + \frac{1}{n\epsilon_n^{1+2j}} \mathsf{V}_j^{\mathrm{BR}}(\mathsf{x}),$$

where the bias constant is

$$\mathsf{B}_j^{\mathrm{BR}}(\mathsf{x}) = \mathcal{D}_{\underline{\mathfrak{s}}+1}(\mathsf{x}) \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathrm{BR}}(k) c_k^{\underline{\mathfrak{s}}+2}$$

and the variance constant is

$$\mathsf{V}_j^{\mathrm{BR}}(\mathsf{x}) = \sum_{k=1}^{\underline{\mathfrak{s}}+1} \sum_{l=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathrm{BR}}(k) \lambda_j^{\mathrm{BR}}(l) \mathcal{C}_{\mathsf{x}}(c_k, c_l);$$

for details, see Section SA.1.7 in the Supplementary Material (Cattaneo, Jansson and Nagasawa (2024)). Assuming $\mathsf{B}_j^{\mathrm{BR}}(\mathsf{x}) \neq 0$, the approximate MSE is minimized by

$$\epsilon_{j,n}^{\mathrm{BR}}(\mathsf{x}) = \left(\frac{1+2j}{2(\underline{\mathfrak{s}}+1-j)} \frac{\mathsf{V}_j^{\mathrm{BR}}(\mathsf{x})}{\mathsf{B}_j^{\mathrm{BR}}(\mathsf{x})^2} \right)^{1/(3+2\underline{\mathfrak{s}})} n^{-1/(3+2\underline{\mathfrak{s}})},$$

a feasible version of which can be constructed by replacing $\mathcal{D}_{\underline{\mathfrak{s}}+1}(\mathsf{x})$ and $\mathcal{C}_{\mathsf{x}}$ with estimators in the expressions for $\mathsf{B}_j^{\mathrm{BR}}(\mathsf{x})$ and $\mathsf{V}_j^{\mathrm{BR}}(\mathsf{x})$, respectively.

4.2. *Bootstrapping.* Suppose inference is to be based on a random sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ from the distribution of some $\mathbf{Z}$. In all examples of which we are aware, the bootstrap parts of Assumption B are satisfied when $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ is given by the nonparametric bootstrap analog of $(\widehat{\Gamma}_n, \widehat{\Phi}_n, \widehat{u}_n)$. Nevertheless, computationally simpler alternatives are often available and in what follows we will present one such alternative. To motivate our proposal, it is instructive to begin by revisiting Example 1.

EXAMPLE 3 (Monotone Density Estimation, continued). In this example, $\mathbf{Z} = X$. Moreover, defining $\gamma_0(x; \mathbf{z}) = \mathbb{1}(\mathbf{z} \leq x)$ and $\phi_0(x; \mathbf{z}) = x$, we have the representations

$$\Gamma_0(x) = \mathbb{E}[\gamma_0(x; \mathbf{Z})] \quad \text{and} \quad \Phi_0(x) = \mathbb{E}[\phi_0(x; \mathbf{Z})],$$

and the estimators $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$ are linear in the sense that they are of the form

$$\widehat{\Gamma}_n(x) = \frac{1}{n} \sum_{i=1}^n \gamma_0(x; \mathbf{Z}_i) \quad \text{and} \quad \widehat{\Phi}_n(x) = \frac{1}{n} \sum_{i=1}^n \phi_0(x; \mathbf{Z}_i),$$

respectively. Finally, $\widehat{u}_n = \max_{1 \leq i \leq n} \mathbf{Z}_i \vee \mathsf{x}$.

Letting $\mathbf{Z}_{1,n}^*, \ldots, \mathbf{Z}_{n,n}^*$ be a random sample from the empirical distribution of $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, $\widehat{u}_n^* = \max_{1 \leq i \leq n} \mathbf{Z}_{i,n}^* \vee \mathsf{x}$ is the nonparametric bootstrap analog of $\widehat{u}_n$, while

$$\widehat{\Gamma}_n^*(x) = \frac{1}{n} \sum_{i=1}^n \gamma_0(x; \mathbf{Z}_{i,n}^*) \quad \text{and} \quad \widehat{\Phi}_n^*(x) = \frac{1}{n} \sum_{i=1}^n \phi_0(x; \mathbf{Z}_{i,n}^*)$$

are the nonparametric bootstrap analogs of $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$, respectively.

In the case of $\widehat{u}_n$, the alternative bootstrap analog $\widehat{u}_n^* = \widehat{u}_n$ is computationally trivial and automatically satisfies the bootstrap parts of (B5) and (B6). As for $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$, their nonparametric bootstrap analogs admit the weighted bootstrap representations

$$\widehat{\Gamma}_n^*(x) = \frac{1}{n} \sum_{i=1}^n W_{i,n} \gamma_0(x; \mathbf{Z}_i) \quad \text{and} \quad \widehat{\Phi}_n^*(x) = \frac{1}{n} \sum_{i=1}^n W_{i,n} \phi_0(x; \mathbf{Z}_i),$$

where, conditionally on $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, $(W_{1,n}, \ldots, W_{n,n})$ is multinomial distributed with probabilities $(n^{-1}, \ldots, n^{-1})$ and number of trials n. (In turn, the weighted bootstrap interpretation of the nonparametric bootstrap analog of $(\widehat{\Gamma}_n, \widehat{\Phi}_n)$ in this example is interesting partly because it can be used to embed the nonparametric bootstrap in a class of bootstraps also containing the Bayesian bootstrap and the wild bootstrap.)

Looking beyond Example 1, finding a computationally trivial $\widehat{u}_n^*$ satisfying the bootstrap part of (B5) and (B6) is usually straightforward. On the other hand, although a weighted bootstrap interpretation of the nonparametric bootstrap version of the estimators $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$ is available whenever they are linear (e.g., in Example 2), there is no shortage of examples for which linearity does not hold. Nevertheless, the weighted bootstrap representation of the nonparametric bootstrap in Example 1 turns out be useful for our purposes, as it is suggestive of computationally attractive alternatives to the nonparametric bootstrap in more complicated examples.

When the nonbootstrap part of (B1) holds, the estimators $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$ are typically asymptotically linear in the sense that they admit (possibly) unknown functions γ_0 and ϕ_0 (satisfying $\Gamma_0(x) = \mathbb{E}[\gamma_0(x; \mathbf{Z})]$ and $\Phi_0(x) = \mathbb{E}[\phi_0(x; \mathbf{Z})]$, respectively) for which the approximations

$$\widehat{\Gamma}_n(x) \approx \bar{\Gamma}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \gamma_0(x; \mathbf{Z}_i) \quad \text{and} \quad \widehat{\Phi}_n(x) \approx \bar{\Phi}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \phi_0(x; \mathbf{Z}_i)$$

are suitably accurate. Assuming also that γ_0 and ϕ_0 admit sufficiently well behaved estimators $\widehat{\gamma}_n$ and $\widehat{\phi}_n$ (say), it then stands to reason that the salient properties of $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$ are well approximated by those of the easy-to-compute exchangeable bootstrap-type pair

$$\widehat{\Gamma}_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} W_{i,n} \widehat{\gamma}_n(x; \mathbf{Z}_i) \quad \text{and} \quad \widehat{\Phi}_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} W_{i,n} \widehat{\phi}_n(x; \mathbf{Z}_i),$$

where $W_{1,n}, \ldots, W_{n,n}$ denote exchangeable random variable (independent of $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$).

To give a precise statement, let

$$\psi_x(v; \mathbf{z}) = \gamma_0(x + v; \mathbf{z}) - \gamma_0(x; \mathbf{z}) - \theta_0(x) [\phi_0(x + v; \mathbf{z}) - \phi_0(x; \mathbf{z})]$$

and define

$$\check{\Gamma}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\gamma}_n(x; \mathbf{Z}_i) \quad \text{and} \quad \check{\Phi}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\phi}_n(x; \mathbf{Z}_i).$$

In addition, for any function class $\mathfrak{F}$, let $N_U(\varepsilon, \mathfrak{F})$ denote the associated uniform covering numbers relative to L_2; that is, for any $\varepsilon > 0$, let

$$N_U(\varepsilon, \mathfrak{F}) = \sup_Q N(\varepsilon \|\bar{F}\|_{Q,2}, \mathfrak{F}, L_2(Q)),$$

where $\bar{F}$ is the minimal envelope function of $\mathfrak{F}$, $\|\cdot\|_{Q,2}$ is the $L_2(Q)$ norm, $N(\cdot)$ is the covering number, and the supremum is over all discrete probability measure Q with $\|\bar{F}\|_{Q,2} > 0$.

ASSUMPTION E. For some $\delta > 0$, the following are satisfied:

(E1) $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, are independent and identically distributed.

(E2) $W_{1,n}, \ldots, W_{n,n}$ are exchangeable random variables independent of $\mathbf{Z}_1, \ldots, \mathbf{Z}_n, \widehat{\gamma}_n$, and $\widehat{\phi}_n$. In addition, for some $\mathfrak{r} > (4\mathfrak{q} + 2)/(2\mathfrak{q} - 1)$,

$$\frac{1}{n} \sum_{i=1}^{n} W_{i,n} = 1, \quad \frac{1}{n} \sum_{i=1}^{n} (W_{i,n} - 1)^2 \to_{\mathbb{P}} 1, \quad \text{and} \quad \mathbb{E}[|W_{1,n}|^{\mathfrak{r}}] = O(1).$$

(E3) $\sup_{x \in I} |\widehat{\Gamma}_n(x) - \bar{\Gamma}_n(x)| = o_\mathbb{P}(1)$ and

$$\frac{1}{n} \sum_{i=1}^n \sup_{x \in I} |\widehat{\gamma}_n(x; \mathbf{Z}_i) - \gamma_0(x; \mathbf{Z}_i)|^2 = o_\mathbb{P}(1).$$

Also,

$$\frac{a_n}{n} \sum_{i=1}^n \sup_{x \in I_x^\delta} |\widehat{\gamma}_n(x; \mathbf{Z}_i) - \gamma_0(x; \mathbf{Z}_i)|^2 = o_\mathbb{P}(1),$$

and there exist $\beta_\gamma \in [1/2, \mathfrak{q} + 1)$ and random variables $A_{\gamma,n} = o_\mathbb{P}(1)$ and $B_{\gamma,n} = o_\mathbb{P}(a_n^{\beta_\gamma})$ such that, for $V \in (0, 2\delta]$,

$$\sqrt{na_n} \sup_{|v| \le V} |\widehat{\Gamma}_n(\mathsf{x} + v) - \widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x} + v) + \bar{\Gamma}_n(\mathsf{x})| \le A_{\gamma,n} + V^{\beta_\gamma} B_{\gamma,n},$$

$$\sqrt{na_n} \sup_{|v| \le V} |\check{\Gamma}_n(\mathsf{x} + v) - \check{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x} + v) + \bar{\Gamma}_n(\mathsf{x})| \le A_{\gamma,n} + V^{\beta_\gamma} B_{\gamma,n}.$$

In addition, for some $\rho_\gamma \in (0, 2)$,

$$\limsup_{\varepsilon \downarrow 0} \frac{\log N_U(\varepsilon, \mathfrak{F}_\gamma)}{\varepsilon^{-\rho_\gamma}} < \infty, \qquad \mathbb{E}[\bar{F}_\gamma(\mathbf{Z})^2] < \infty, \qquad \limsup_{\varepsilon \downarrow 0} \frac{\log N_U(\varepsilon, \widehat{\mathfrak{F}}_{\gamma,n})}{\varepsilon^{-\rho_\gamma}} = O_\mathbb{P}(1),$$

where $\mathfrak{F}_\gamma = \{\gamma_0(x; \cdot) : x \in I\}$, $\bar{F}_\gamma$ is its minimal envelope, and $\widehat{\mathfrak{F}}_{\gamma,n} = \{\widehat{\gamma}_n(x; \cdot) : x \in I\}$.
    Also,

$$\limsup_{\eta \downarrow 0} \frac{\mathbb{E}[\bar{D}_\gamma^\eta(\mathbf{Z})^2 + \bar{D}_\gamma^\eta(\mathbf{Z})^4]}{\eta} < \infty,$$

where $\bar{D}_\gamma^\eta$ is the minimal envelope of $\{\gamma_0(x; \cdot) - \gamma_0(\mathsf{x}; \cdot) : x \in I_\mathsf{x}^\eta\}$.
    (E4) $\sup_{x \in I} |\widehat{\Phi}_n(x) - \bar{\Phi}_n(x)| = o_\mathbb{P}(1)$ and

$$\frac{1}{n} \sum_{i=1}^n \sup_{x \in I} |\widehat{\phi}_n(x; \mathbf{Z}_i) - \phi_0(x; \mathbf{Z}_i)|^2 = o_\mathbb{P}(1).$$

Also, $a_n |\widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x})| = o_\mathbb{P}(1)$, $a_n |\check{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x})| = o_\mathbb{P}(1)$,

$$\frac{a_n}{n} \sum_{i=1}^n \sup_{x \in I_x^\delta} |\widehat{\phi}_n(x; \mathbf{Z}_i) - \phi_0(x; \mathbf{Z}_i)|^2 = o_\mathbb{P}(1),$$

and there exist $\beta_\phi \in [1/2, \mathfrak{q}]$ and random variables $A_{\phi,n} = o_\mathbb{P}(1)$ and $B_{\phi,n} = o_\mathbb{P}(a_n^{\beta_\phi})$ such that, for $V \in (0, 2\delta]$,

$$\sqrt{na_n} \sup_{|v| \le V} |\widehat{\Phi}_n(\mathsf{x} + v) - \widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x} + v) + \bar{\Phi}_n(\mathsf{x})| \le A_{\phi,n} + V^{\beta_\phi} B_{\phi,n},$$

$$\sqrt{na_n} \sup_{|v| \le V} |\check{\Phi}_n(\mathsf{x} + v) - \check{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x} + v) + \bar{\Phi}_n(\mathsf{x})| \le A_{\phi,n} + V^{\beta_\phi} B_{\phi,n}.$$

In addition, for some $\rho_\phi \in (0, 2)$,

$$\limsup_{\varepsilon \downarrow 0} \frac{\log N_U(\varepsilon, \mathfrak{F}_\phi)}{\varepsilon^{-\rho_\phi}} < \infty, \qquad \mathbb{E}[\bar{F}_\phi(\mathbf{Z})^2] < \infty, \qquad \limsup_{\varepsilon \downarrow 0} \frac{\log N_U(\eta, \widehat{\mathfrak{F}}_{\phi,n})}{\varepsilon^{-\rho_\phi}} = O_\mathbb{P}(1),$$

where $\mathfrak{F}_\phi = \{\phi_0(x; \cdot) : x \in I\}$, $\bar{F}_\phi$ is its minimal envelope, and $\widehat{\mathfrak{F}}_{\phi,n} = \{\widehat{\phi}_n(x; \cdot) : x \in I\}$.

Also,

$$\limsup_{\eta \downarrow 0} \frac{\mathbb{E}[\bar{D}_\phi^\eta(\mathbf{Z})^2 + \bar{D}_\phi^\eta(\mathbf{Z})^4]}{\eta} < \infty,$$

where $\bar{D}_\phi^\eta$ is the minimal envelope of $\{\phi_0(x; \cdot) - \phi_0(\mathsf{x}; \cdot) : x \in I_\mathsf{x}^\eta\}$.

(E5) For every $\eta_n > 0$ with $a_n \eta_n = O(1)$,

$$\sup_{v, v' \in [-\eta_n, \eta_n]} \frac{\mathbb{E}[|\psi_\mathsf{x}(v; \mathbf{Z}) - \psi_\mathsf{x}(v'; \mathbf{Z})|]}{|v - v'|} = O(1)$$

and, for all $v, u \in \mathbb{R}$, and for some $\mathcal{C}_\mathsf{x}$ (as defined in (B1)),

$$\frac{\mathbb{E}[\psi_\mathsf{x}(\eta_n v; \mathbf{Z})\psi_\mathsf{x}(\eta_n u; \mathbf{Z})]}{\eta_n} \to \mathcal{C}_\mathsf{x}(v, u).$$

LEMMA 3. *Suppose Assumptions* A *and* E *are satisfied. Then* (B1)-(B4) *are satisfied.*

5. Examples. We apply our main results to two distinct sets of examples, both previously analyzed in Westling and Carone (2020) and Westling, Gilbert and Carone (2020), and references therein. In the Supplementary Material, we also consider two other set of examples, namely monotone hazard estimation (Huang and Wellner (1995)) and monotone distribution estimation (van der Vaart and van der Laan (2006)). All the proofs are found in the Supplementary Material.

For future reference, we collect the conditions on the bootstrap weights in an assumption.

ASSUMPTION BW. The bootstrap weights $W_{1,n}, \ldots, W_{n,n}$ are exchangeable, independent of the random sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, and satisfy, for some $\mathfrak{r} > 6$,

$$\frac{1}{n}\sum_{i=1}^n W_{i,n} = 1, \qquad \frac{1}{n}\sum_{i=1}^n (W_{i,n} - 1)^2 \to_\mathbb{P} 1, \quad \text{and} \quad \mathbb{E}[|W_{1,n}|^\mathfrak{r}] = O(1).$$

5.1. *Monotone density estimation.* Consider the problem of estimating the density of a nonnegative, continuously distributed random variable, possibly with censoring. Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be i.i.d. copies of $\mathbf{Z} = (\check{X}, \Delta)'$, with $\check{X} = X \wedge C$ and $\Delta = \mathbb{1}(X \le C)$. Assuming that f_0, the density of X, (exists and) is nondecreasing on $I = [0, u_0]$ where u_0 is some point in the support of X, the parameter of interest is $\theta_0(\mathsf{x})$ for some $\mathsf{x} \in (0, u_0)$, where $\theta_0 = f_0$. This setup generalizes Example 1. We take $\Phi_0(x) = x$ and therefore have $\Gamma_0 = F_0$, where F_0 is the cdf of X.

The canonical case of no censoring (i.e., $\mathbb{P}[C \ge X] = 1$) has been considered in Example 1. Assumptions under which this example is covered by our general theory is given in the following result.

COROLLARY 1. *Let u_0 be the supremum of the support of X. Suppose x is in the interior of I, θ_0 satisfies (A2), and Assumption BW holds. Then, Assumptions* A, B, *and* D *hold with*

$$\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \le x), \qquad \widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\mathbb{1}(X_i \le x),$$

$$\widehat{\Phi}_n(x) = \widehat{\Phi}_n^*(x) = x, \qquad \widehat{u}_n = \widehat{u}_n^* = \max_{1 \le i \le n} X_i \vee \mathsf{x},$$

$$\mathcal{C}_\mathsf{x}(s, t) = f_0(\mathsf{x})(|s| \wedge |t|)\mathbb{1}(\text{sign}(s) = \text{sign}(t)), \qquad \mathcal{D}_\mathfrak{q}(\mathsf{x}) = \frac{\partial^\mathfrak{q} f_0(\mathsf{x})}{(\mathfrak{q} + 1)!}.$$

Under the assumptions of Corollary 1, it follows from Theorem 1 that the bootstrap consistency result (8) holds for any $\widetilde{M}_{\mathsf{x},n}$ satisfying Assumption C. In particular, any pointwise consistent estimator of $\partial^{\mathsf{q}} f_0(\mathsf{x})$ could be used to estimate $\mathcal{D}_{\mathsf{q}}(\mathsf{x})$. Alternatively, since Assumption D also holds, $\widetilde{M}_{\mathsf{x},n}$ in (10) with $\widetilde{\mathcal{D}}_{\mathsf{q},n} \in \{\widetilde{\mathcal{D}}_{\mathsf{q},n}^{\mathrm{MA}}, \widetilde{\mathcal{D}}_{\mathsf{q},n}^{\mathrm{FD}}, \widetilde{\mathcal{D}}_{\mathsf{q},n}^{\mathrm{BR}}\}$ can also be used, provided that $\epsilon_n \to 0$ and $n\epsilon_n^{1+2\mathsf{q}} \to \infty$.

Suppose now that censoring occurs completely at random; that is, suppose $X \perp\!\!\!\perp C$ (e.g., Huang and Wellner (1995), and references therein). Let $\widehat{S}_n$ and $\widehat{G}_n$ be the Kaplan-Meier estimators of the survival functions $S_0(\cdot) = \mathbb{P}[X > \cdot]$ and $G_0(\cdot) = \mathbb{P}[C > \cdot]$, and define $\widehat{F}_n = 1 - \widehat{S}_n$ and $\widehat{\Lambda}_n(x) = \int_0^x \widehat{S}_n(u)^{-1} d\widehat{F}_n(u)$. In anticipation of the following result, we note that $\Gamma_0(x) = F_0(x) = \mathbb{E}[\gamma_0(x; \mathbf{Z})]$, where

$$\gamma_0(x; \mathbf{Z}) = F_0(x) + S_0(x)\left[\frac{\Delta \mathbb{1}(\check{X} \le x)}{S_0(\check{X})G_0(\check{X})} - \int_0^{\check{X} \wedge x} \frac{\Lambda_0(du)}{S_0(u)G_0(u)}\right],$$

with $\Lambda_0(x) = \int_0^x S_0(u)^{-1} dF_0(u)$.

COROLLARY 2. *Let u_0 be an interior point in the support of X. Suppose x is in the interior of I, θ_0 satisfies (A2), $C \perp\!\!\!\perp X$, C is absolutely continuous on I with bounded density, and $S_0(u_0)G_0(u_0) > 0$. Then, Assumptions A, B, and D hold with*

$$\widehat{\Gamma}_n(x) = \widehat{F}_n(x), \qquad \widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\gamma}_n(x; \mathbf{Z}_i),$$

$$\widehat{\gamma}_n(x; \mathbf{Z}) = \widehat{F}_n(x) + \widehat{S}_n(x)\left[\frac{\Delta \mathbb{1}(\check{X} \le x)}{\widehat{S}_n(\check{X})\widehat{G}_n(\check{X})} - \int_0^{\check{X} \wedge x} \frac{d\widehat{\Lambda}_n(u)}{\widehat{S}_n(u)\widehat{G}_n(u)}\right],$$

$$\widehat{\Phi}_n(x) = \widehat{\Phi}_n^*(x) = x, \qquad \widehat{u}_n = \widehat{u}_n^* = u_0,$$

$$\mathcal{C}_{\mathsf{x}}(s, t) = \frac{f_0(\mathsf{x})}{G_0(\mathsf{x})}\left(|s| \wedge |t|\right)\mathbb{1}\left(\mathrm{sign}(s) = \mathrm{sign}(t)\right), \qquad \mathcal{D}_{\mathsf{q}}(\mathsf{x}) = \frac{\partial^{\mathsf{q}} f_0(\mathsf{x})}{(\mathsf{q}+1)!}.$$

Theorem 1 implies that inference based on our proposed bootstrap-assisted method is asymptotically valid for this example as well. Additionally, Assumption D also holds under the above conditions, and we can use one of the numerical derivative estimators discussed in Section 4.1.

In the Supplementary Material, we also consider estimation of monotone density under conditionally independent right-censoring, where covariates are present. We do not discuss this example here because it involves additional notation and technicalities.

5.2. *Monotone regression estimation.* Consider now the problem of regression estimation, possibly with covariate adjustment. We assume that $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are i.i.d. copies of $\mathbf{Z} = (Y, X, \mathbf{A}')'$, where X is continuous on its support I with density f_0 and cdf F_0. Defining $\mu_0(x, \mathbf{a}) = \mathbb{E}[Y|X = x, \mathbf{A} = \mathbf{a}]$, the parameter of interest is $\theta_0(\mathsf{x})$ for some x in the interior of I, where $\theta_0(\cdot) = \mathbb{E}[\mu_0(\cdot, \mathbf{A})]$ is assumed to be nondecreasing on I. (If there are no covariates $\mathbf{A}$, then $\theta_0(\cdot) = \mu_0(\cdot) = \mathbb{E}[Y|X = \cdot]$.) We take $\Phi_0 = F_0$ and set $u_0 = 1$.

The classical monotone regression estimator has been considered in Example 2. Assumptions under which this example is covered by our general theory is given in the following result. To state it, let $\varepsilon = Y - \mathbb{E}[Y|X]$ and define $\sigma_0^2(x) = \mathbb{E}[\varepsilon^2|X = x]$.

COROLLARY 3. *Suppose that x is in the interior of I, θ_0 satisfies (A2), Φ_0 satisfies (A3), and Assumption BW holds. If $\sigma_0^2(\mathsf{x})$ is continuous and positive at x, $\mathbb{E}[\varepsilon^2] < \infty$, and if*

$\sup_{x \in I_x^\delta} \mathbb{E}[\varepsilon^4 | X = x] < \infty$ *for some* $\delta > 0$, *then Assumptions* A, B, *and* D *hold with*

$$\widehat{\Gamma}_n(x) = \frac{1}{n} \sum_{i=1}^n Y_i \mathbb{1}(X_i \le x), \qquad \widehat{\Gamma}_n^*(x) = \frac{1}{n} \sum_{i=1}^n W_{i,n} Y_i \mathbb{1}(X_i \le x),$$

$$\widehat{\Phi}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \le x), \qquad \widehat{\Phi}_n^*(x) = \frac{1}{n} \sum_{i=1}^n W_{i,n} \mathbb{1}(X_i \le x), \qquad \widehat{u}_n = \widehat{u}_n^* = 1,$$

$$\mathcal{C}_x(s, t) = f_0(x) \sigma_0^2(x) \big(|s| \wedge |t|\big) \mathbb{1}\big(\text{sign}(s) = \text{sign}(t)\big), \qquad \mathcal{D}_q(x) = \frac{f_0(x) \partial^q \mu_0(x)}{(q+1)!}.$$

Under the assumptions of Corollary 3, it follows from Theorem 1 that the bootstrap consistency result (8) holds for any $\widetilde{M}_{x,n}$ satisfying Assumption C. In this example, any pointwise consistent estimators of $f_0(x)$ and $\partial^q \mu_0(x)$ could be used to estimate $\mathcal{D}_q(x)$. Alternatively, instead of using two distinct estimators, we can use the numerical derivative-type estimators since Assumption D also holds.

Next, consider the case of monotone regression estimation with covariate-adjustment. Introducing covariates $\mathbf{A}$, let $f_{X|\mathbf{A}}(\cdot | \mathbf{a})$ denote the conditional Lebesgue density of X given $\mathbf{A} = \mathbf{a}$ and define $g_0(x, \mathbf{a}) = f_{X|\mathbf{A}}(x | \mathbf{a}) / f_0(x)$. Noting that $\Gamma_0(x) = \mathbb{E}[\gamma_0(x; \mathbf{Z})]$, where

$$\gamma_0(x; \mathbf{Z}) = \mathbb{1}(X \le x) \bigg[\frac{Y - \mu_0(X, \mathbf{A})}{g_0(X, \mathbf{A})} - \theta_0(X) \bigg],$$

we assume that $\widehat{\Gamma}_n$ is of the form $\widehat{\Gamma}_n(x) = n^{-1} \sum_{i=1}^n \widehat{\gamma}_n(x; \mathbf{Z}_i)$, where

$$\widehat{\gamma}_n(x; \mathbf{Z}) = \mathbb{1}(X \le x) \bigg[\frac{Y - \widehat{\mu}_n(X, \mathbf{A})}{\widehat{g}_n(X, \mathbf{A})} + \frac{1}{n} \sum_{j=1}^n \widehat{\mu}_n(X, \mathbf{A}_j) \bigg],$$

with $\widehat{\mu}_n$ and $\widehat{g}_n$ being preliminary estimators of μ_0 and g_0 based on $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. As discussed by Westling, Gilbert and Carone (2020), there are a number of candidates for $\widehat{\mu}_n$ and $\widehat{g}_n$. To accommodate a relatively wide class of estimators, the following assumption employs high-level conditions on $\widehat{\mu}_n$ and $\widehat{g}_n$. To state the assumption, let $\varepsilon = Y - \mathbb{E}[Y | X, \mathbf{A}]$ and define $\sigma_0^2(x, \mathbf{a}) = \mathbb{E}[\varepsilon^2 | X = x, \mathbf{A} = \mathbf{a}]$.

ASSUMPTION MRC. For some $\delta > 0$, the following are satisfied:

(i) $\mathbb{E}[g_0(x, \mathbf{A})^{-1} \sigma_0^2(x, \mathbf{A})] > 0$, $\mathbb{E}[\varepsilon^2] < \infty$, and $\sup_{x \in I_x^\delta} \mathbb{E}[\varepsilon^4 | X = x] < \infty$.
(ii) $f_{X|\mathbf{A}}$ is bounded and g_0 is bounded away from zero.
(iii) There exist random variables $A_n = o_{\mathbb{P}}(1)$ and $B_n = O_{\mathbb{P}}(a_n^{1/2})$ such that

$$\sqrt{na_n} \sup_{|v| \le V} \big| \widehat{\Gamma}_n(x + v) - \widehat{\Gamma}_n(x) - \bar{\Gamma}_n(x + v) + \bar{\Gamma}_n(x) \big| \le A_n + V B_n, \qquad V \in (0, 2\delta].$$

In addition,

$$\frac{a_n}{n} \sum_{i=1}^n \big| \widehat{\mu}_n(X_i, \mathbf{A}_i) - \mu_0(X_i, \mathbf{A}_i) \big|^2 = o_{\mathbb{P}}(1),$$

$$\frac{a_n}{n^2} \sum_{i=1}^n \sum_{j=1}^n \big| \widehat{\mu}_n(X_i, \mathbf{A}_j) - \mu_0(X_i, \mathbf{A}_j) \big|^2 = o_{\mathbb{P}}(1),$$

and

$$\frac{a_n}{n} \sum_{i=1}^n \varepsilon_i^2 \big| \widehat{g}_n(X_i, \mathbf{A}_i) - g_0(X_i, \mathbf{A}_i) \big|^2 = o_{\mathbb{P}}(1).$$

(iv) $\mathbb{E}[\bar{\mu}(\mathbf{A})^2] < \infty$, where

$$\bar{\mu}(\mathbf{A}) = \sup_{|x-x'| \leq \delta} \frac{|\mu_0(x, \mathbf{A}) - \mu_0(x', \mathbf{A})|}{|x - x'|}.$$

(v) $\mathbb{E}[\bar{\sigma}^2(\mathbf{A})] < \infty$, where, for some function ω with $\lim_{\eta \downarrow 0} \omega(\eta) = 0$,

$$\left| \frac{\sigma_0^2(x, \mathbf{A}) f_0(x)}{g_0(x, \mathbf{A})} - \frac{\sigma_0^2(\mathsf{x}, \mathbf{A}) f_0(\mathsf{x})}{g_0(\mathsf{x}, \mathbf{A})} \right| \leq \omega(|x - \mathsf{x}|) \bar{\sigma}^2(\mathbf{A}), \quad x \in I_{\mathsf{x}}^{\delta}.$$

Part (iii) of the assumption contains high-level conditions. In the Supplementary Material, we provide primitive sufficient conditions using nonparametric estimators based on sample splitting.

COROLLARY 4. *Suppose that* x *is in the interior of* I, θ_0 *satisfies* (A2), Φ_0 *satisfies* (A3), *and Assumption BW holds. If Assumption MRC holds, then Assumptions* A *and* B *hold with*

$$\widehat{\Gamma}_n^*(x) = \frac{1}{n} \sum_{i=1}^n W_{i,n} \widehat{\gamma}_n(x; \mathbf{Z}_i),$$

$$\widehat{\Phi}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x), \qquad \widehat{\Phi}_n^*(x) = \frac{1}{n} \sum_{i=1}^n W_{i,n} \mathbb{1}(X_i \leq x), \qquad \widehat{u}_n = \widehat{u}_n^* = 1,$$

$$\mathcal{C}_{\mathsf{x}}(s, t) = f_0(\mathsf{x}) \mathbb{E}\left[\frac{\sigma_0^2(\mathsf{x}, \mathbf{A})}{g_0(\mathsf{x}, \mathbf{A})} \right] (|s| \wedge |t|) \mathbb{1}(\text{sign}(s) = \text{sign}(t)), \qquad \mathcal{D}_{\mathsf{q}}(\mathsf{x}) = \frac{f_0(\mathsf{x}) \partial^{\mathsf{q}} \theta_0(\mathsf{x})}{(\mathsf{q} + 1)!}.$$

As a consequence, if $\widetilde{M}_{\mathsf{x},n}$ satisfies Assumption C, then the "reshaped" bootstrap estimator $\widetilde{\theta}_n^*(\mathsf{x}) = \partial_- \mathsf{GCM}_{[0,1]}(\widetilde{\Gamma}_n^* \circ \widehat{\Phi}_n^{*-}) \circ \widehat{\Phi}_n^*(\mathsf{x})$ gives a bootstrap-assisted distributional approximation satisfying (8). In this example, the covariance kernel contains $\mathbb{E}[g_0(\mathsf{x}, \mathbf{A})^{-1} \sigma_0^2(\mathsf{x}, \mathbf{A})]$ and the plug-in procedure discussed in Section 3 would involve estimating this nuisance parameter, which in turn requires estimation of $\sigma_0^2(\mathsf{x}, \cdot)$. When $\mathbf{A}$ is of moderate dimension, this would introduce a challenging estimation problem on top of estimating μ_0 and g_0. Our bootstrap-assisted procedure circumvents estimation of this additional nuisance parameter.

6. Simulations. We consider the canonical case of monotone regression (i.e., Example 2). We estimate the nondecreasing regression function $\theta_0(\cdot) = \mathbb{E}[Y|X = \cdot]$ at an interior point x using a random sample of observations. Three distinct data generating processes (DGPs) are considered. For each DGP, we let X be a uniform $(0, 1)$ random variable and we let $Y = \theta_0(X) + \sigma_0(X)\widetilde{\varepsilon}$, where $\widetilde{\varepsilon}$ is a standard normal random variable independent of X. DGP 1 sets $\theta_0(x) = 2\exp(x - 0.5)$ and $\sigma_0 = 1$, DGP 2 sets $\theta_0(x) = 2\exp(x - 0.5)$ and $\sigma_0 = \exp(x)$, and DGP 3 sets $\theta_0(x) = 24\exp(X - 0.5) - 24(X - 0.5) - 12(X - 0.5)^2$ and $\sigma_0 = 0.1$. In all three cases, we set $\mathsf{x} = 0.5$. The second DGP exhibits heteroskedastic regression errors, and the third DGP features a regression function whose first derivative equals zero at the evaluation point x; that is, DGP 3 exhibits a degree of degeneracy and if inference is conducted without knowledge of the fact that $\mathsf{q} = 3$, then this feature makes the inference problem more challenging.

The Monte Carlo experiment employs a sample size $n = 1000$ with 2000 bootstrap replications and 4000 simulations, and compares four types of bootstrap-based inference procedures: the standard nonparametric bootstrap, m-out-of-n bootstrap, the "plug-in method" which simulates the limit law with estimated nuisance parameters, and our proposed bootstrap-assisted inference method implemented using the bias-reduced numerical derivative estimator discussed in Section 4.1. For the plug-in method, we estimate the covariance

TABLE 1
Simulations, Monotone Regression Estimator, 95% Confidence Intervals

	DGP 1				DGP 2				DGP 3			
	$\tilde{\mathcal{D}}_{1,n}$	$\tilde{\mathcal{D}}_{3,n}$	Coverage	Length	$\tilde{\mathcal{D}}_{1,n}$	$\tilde{\mathcal{D}}_{3,n}$	Coverage	Length	$\tilde{\mathcal{D}}_{1,n}$	$\tilde{\mathcal{D}}_{3,n}$	Coverage	Length
Standard			0.828	0.373			0.838	0.519			0.912	0.029
m-out-of-n												
$m = \lceil n^{1/2} \rceil$			0.896	0.413			0.915	0.583			0.939	0.031
$m = \lceil n^{2/3} \rceil$			0.868	0.400			0.884	0.557			0.928	0.029
$m = \lceil n^{4/5} \rceil$			0.851	0.393			0.864	0.545			0.918	0.029
Plug-in	1.053	0.000	0.950	0.399	1.052	0.000	0.946	0.553	0.000	1.003	0.936	0.028
Reshaped												
Oracle	1.000	0.000	0.941	0.395	1.000	0.000	0.951	0.550	0.000	1.000	0.946	0.029
ND known q	1.053	0.000	0.949	0.398	1.052	0.000	0.945	0.547	0.000	1.003	0.936	0.028
ND robust	1.053	0.599	0.950	0.401	1.052	0.938	0.954	0.559	0.014	1.003	0.960	0.030

Notes: (i) Panel *Standard* refers to standard nonparametric bootstrap, Panel *m-out-of-n* refers to *m*-out-of-*n* non-parametric bootstrap with subsample *m*, Panel *Plug-in* refers to the plug-in method, Panel *Reshaped* refers to our proposed bootstrap-assisted procedure.

(ii) Columns "$\tilde{\mathcal{D}}_{1,n}$" and "$\tilde{\mathcal{D}}_{3,n}$" report the averages of the estimated $\mathcal{D}_1$, $\mathcal{D}_3$ across simulations, and Columns "Coverage" and "Length" report empirical coverage and average length of bootstrap-based 95% percentile confidence intervals, respectively.

(iii) "Oracle" corresponds to the infeasible version of our proposed procedure using the true value of $\mathcal{D}_q$, "ND known q" corresponds to our proposed procedure using the bias-reduced numerical derivative estimator with a correct specification of q, and "ND robust" corresponds to our proposed procedure only assuming q $\in \{1, 3\}$. The step size choice for the numerical derivative estimator is described in the Supplementary Material.

(iv) The sample size is 1000, the number of bootstrap iterations is 2000, and the number of Monte Carlo simulations is 4000.

kernel and the nuisance parameter $\partial \Phi_0(\mathsf{x})$ using local linear kernel estimators, and we estimate the mean function by the same numerical derivative estimator as for our bootstrap-assisted procedure. For *m*-out-of-*n* bootstrap and the plug-in method, researchers need to specify the characteristic exponent q, and we use the true (infeasible) value. For the numerical derivative estimator, we developed a rule of thumb for the step size ϵ_n to operationalize the procedure. For further details, see Section SA.6 of Cattaneo, Jansson and Nagasawa (2024). For our proposed method, we report results for three implementations: (i) (infeasible) procedure using the true value of $\mathcal{D}_0(\mathsf{x})$ (referred to as "oracle"), (ii) implementation using the numerical derivative estimator with a correct specification of q (referred to as "known q"), and (iii) "robust" implementation only assuming q $\in \{1, 3\}$ (referred to as "robust").

Table 1 presents the numerical results. We report empirical coverage for nominal 95% confidence intervals and their average interval length. For all of the DGPs considered, our proposed bootstrap-assisted inference method leads to confidence intervals with excellent empirical coverage and average interval length. The infeasible "oracle" procedure attains empirical coverage very close to the nominal 95%, which aligns with our theoretical results, and feasible procedures using the numerical derivative estimators perform almost identical to the infeasible oracle version.

In DGPs 1 and 2, our procedures outperform both the standard nonparametric bootstrap (which is inconsistent) and the *m*-out-of-*n* bootstrap (which is consistent) in empirical coverage and average length. In DGP 3, the *m*-out-of-*n* with the subsample size $m = \lceil n^{1/2} \rceil$ performs comparable to our procedure, but two caveats should be noted. First, the *m*-out-of-*n* bootstrap performance is sensitive to the choice of the subsample size. Therefore, to operationalize the *m*-out-of-*n* bootstrap procedure, one needs to develop a reliable procedure to choose the subsample size. Another caveat, arguably more important in this context, is

that the m-out-of-n bootstrap procedure requires the knowledge of the convergence rate of the estimator. In our simulation, we use the true convergence rate for m-out-of-n bootstrap, but in practice, one needs to assume or estimate from data the convergence rate. Since the convergence rate and the limit distribution of the generalized Grenander-type estimator crucially hinge on unknown q, this feature of the m-out-of-n bootstrap may be unappealing. In contrast, our proposed bootstrap-based procedure only requires specifying an upper bound on q and it automatically adapts to the unknown convergence rate.

The plug-in method performs on par with our method, which is not unexpected because (i) the covariance kernel in this setting takes a relatively simple form and estimating it is not challenging with sample size $n = 1000$ and (ii) we use the true q to simulate the limit law. In other examples where covariates are present, estimating covariance kernels typically involves nonparametric estimation of additional, possibly high-dimensional, nuisance functions. Our proposed method reduces the number of nuisance parameters that need to be estimated. Moreover, the good performance of the plug-in method for DGP 3 crucially depends on correctly specifying q, whereas our "robust" procedure only requires specifying an upper bound on q, and this is another desirable feature of our proposed method.

APPENDIX: TECHNICAL RESULTS AND OMITTED DETAILS

This appendix collects several technical results and details omitted from the main paper to improve exposition. First, in Section A.1 we present a corrected version of the generalization of the switch relation stated by Westling and Carone (2020, Supplement). Second, in Section A.2 we present a lemma that can be used to establish continuity of the cdf of the maximizer of a Gaussian process whose covariance kernel is that of two-sided Brownian motion. Both lemmas are used in the proof of Theorem 1 and may be of independent interest as well. Third, in Section A.3 we establish inconsistency of the plug-in nonparametric bootstrap for a large class of Generalized Grenander-type Estimators, following early work by Kosorok (2008) and Sen, Banerjee and Woodroofe (2010). Finally, in Section A.4 we report omitted formulas from Section 3.1.

A.1. Generalized switch relation. For a real-valued function f defined on a set $X \subset \mathbb{R}$, $\operatorname{argmax}_{x \in X}\{f(x)\}$ denotes the (possibly empty) set of maximizers of f over X. If $\operatorname{argmax}_{x \in X}\{f(x)\}$ is nonempty and contains a largest element, we denote that element by $\max \operatorname{argmax}_{x \in X}\{f(x)\}$.

In their analysis of generalized Grenander-type estimators, Westling and Carone (2020) relied on a generalization of the *switch relation* (Groeneboom (1985)). Their generalized switch relation is given in Lemma 1 of Westling and Carone (2020, Supplement). For the purposes of comparing it with Lemma A.1 below, it is convenient to restate that lemma as follows:

Statement GSR (Westling and Carone (2020)). *Let Φ and Γ be real-valued functions defined on an interval $I \subseteq \mathbb{R}$ and suppose that Φ is nondecreasing and right-continuous. Fix l, u in $\Phi(I)$ with $l < u$ and let $\theta = \partial_- \mathsf{GCM}_{[l,u]}(\Gamma \circ \Phi^-) \circ \Phi$. If I is closed, $\Phi(I) \subseteq [l, u]$, and if Γ and $\Gamma \circ \Phi^-$ are lower semicontinuous, then*

$$(13) \qquad \theta(\mathsf{x}) > t \quad \Longleftrightarrow \quad \sup_{x \in \Phi^-([l,u])} \operatorname{argmax} \{t\Phi(x) - \Gamma(x)\} < \Phi^-(\Phi(\mathsf{x}))$$

for any $t \in \mathbb{R}$ and any $\mathsf{x} \in I$ with $\Phi(\mathsf{x}) \in (l, u)$.

The main purpose of the following example is to show that without further restrictions, the argmax in (13) can be empty and the relation (13) can be violated (if we interpret $\sup \operatorname{argmax}_{x \in X}\{f(x)\}$ as $\inf X$ when $\operatorname{argmax}_{x \in X}\{f(x)\}$ is empty). In other words, Statement GSR can fail to hold.

EXAMPLE A.1. Let $I = [l, u] = [0, 1]$, $\Gamma(x) = \gamma x$ (for some $\gamma \in \mathbb{R}$), and let

$$\Phi(x) = \begin{cases} x & \text{if } 0 \leq x < 1/2, \\ 1 & \text{if } 1/2 \leq x \leq 1. \end{cases}$$

Then

$$t\Phi(x) - \Gamma(x) = \begin{cases} (t - \gamma)x & \text{if } 0 \leq x < 1/2, \\ t - \gamma/2 & \text{if } 1/2 \leq x \leq 1, \end{cases}$$

so $\operatorname{argmax}_{x \in \Phi^-([l,u])}\{t\Phi(x) - \Gamma(x)\}$ is empty when $0 > t > \gamma$.

In particular, if $0 > t > \gamma$ and if $\mathsf{x} \in (0, 1/2)$, then

$$\sup_{x \in \Phi^-([l,u])} \operatorname{argmax} \{t\Phi(x) - \Gamma(x)\} = 0 < \mathsf{x} = \Phi^-(\Phi(\mathsf{x})),$$

whereas $\theta(\mathsf{x}) = \partial_- \mathsf{GCM}_{[l,u]}(\Gamma \circ \Phi^-) \circ \Phi(\mathsf{x}) = \gamma < t$, so (13) is violated.

For future reference, we note that $\Phi(I) = \Phi(I) \cap [l, u] = [0, 1/2) \cup \{1\}$ in this example.

The problems highlighted by the example are attributable to the fact that $\Phi(I) \cap [l, u]$ is not closed. Fortunately, it turns out that one can obtain a result in the spirit of (13) as long as $\Phi(I) \cap [l, u]$ is closed. The following lemma gives a precise statement.

LEMMA A.1 (Generalized Switch Relation). *Let Φ and Γ be real-valued functions defined on an interval $I \subseteq \mathbb{R}$ and suppose that Φ is nondecreasing and right-continuous. Fix $l, u \in \Phi(I)$ with $l < u$ and let $\theta = \partial_- \mathsf{GCM}_{[l,u]}(\Gamma \circ \Phi^-) \circ \Phi$. If $\Phi(I) \cap [l, u]$ is closed, then*

$$(14) \qquad \theta(\mathsf{x}) > t \quad \Longleftrightarrow \quad \max_{x \in \Phi^-([l,u])} \operatorname{argmax} \{t\Phi(x) - \mathsf{LSC}_\Phi(\Gamma)(x)\} < \Phi^-(\Phi(\mathsf{x}))$$

for any $t \in \mathbb{R}$ and any $\mathsf{x} \in I$ with $\Phi(\mathsf{x}) \in (l, u)$, where

$$\mathsf{LSC}_\Phi(\Gamma)(x) = \mathsf{LSC}(\Gamma \circ \Phi^-) \circ \Phi(x),$$

and where $\mathsf{LSC}(\cdot)$ denotes the greatest lower semicontinuous minorant.

The assumptions of the lemma seem mild. In particular, the assumption that $\Phi(I) \cap [l, u]$ is closed is satisfied not only in the examples of Section 5, but in all examples of which we are aware.

The lemma does not assume lower semicontinuity of Γ or $\Gamma \circ \Phi^-$. Instead, the conclusion involves the function $\mathsf{LSC}_\Phi(\Gamma)$. As the notation suggests, $\mathsf{LSC}_\Phi(\cdot) = \mathsf{LSC}(\cdot)$ when Φ is the identity. More generally, $\mathsf{LSC}_\Phi(\cdot) = \mathsf{LSC}(\cdot)$ when Φ is continuous and strictly increasing and whether or not these properties hold it follows from Lemma 4.3 of van der Vaart and van der Laan (2006) that $\mathsf{LSC}_\Phi(\Gamma)$ admits the following characterization:

$$\mathsf{LSC}_\Phi(\Gamma)(x) = \liminf_{y \to \Phi(x)} (\Gamma \circ \Phi^-)(y).$$

A.2. Continuity of argmax. Let $\{\mathbb{G}(v) : v \in \mathbb{R}\}$ be a Gaussian process with $\mathbb{E}[\mathbb{G}(v)] = \mu(v)$ and $\mathbb{C}\mathrm{ov}[\mathbb{G}(v), \mathbb{G}(u)] = C(|v| \wedge |u|)\mathbb{1}(\mathrm{sign}(v) = \mathrm{sign}(u))$ for some $C > 0$, i.e. $\mathbb{G}$ is a scalar multiple of two-sided Brownian motion with mean shift. Under conditions on μ stated below, there exists a unique maximizer of $\mathbb{G}(v)$ over $v \in \mathbb{R}$ with probability one (Kim and Pollard (1990)). We complement this known fact with a result showing that the cdf of $\operatorname{argmax}_{v \in \mathbb{R}}\{\mathbb{G}(v)\}$ is continuous.

LEMMA A.2. *Suppose that for each $v \in \mathbb{R}$, $\lim_{\delta \downarrow 0}[\mu(v + \delta) - \mu(v)]/\sqrt{\delta} = 0$ and that $\limsup_{|v| \to \infty} \mu(v)/|v|^c = -\infty$ for some $c > 1/2$. Then $x \mapsto \mathbb{P}[\arg\max_{v \in \mathbb{R}}\{\mathbb{G}(v)\} \leq x]$ is continuous.*

Under Assumptions A and B of the main text, $-\mathcal{M}_{\mathsf{x}}^{\mathsf{q}} + t\mathcal{L}_{\mathsf{x}}$ satisfies the hypothesis of Lemma A.2 for any $t \in \mathbb{R}$. It therefore follows from the lemma that the function

$$x \mapsto \mathbb{P}\Big[\arg\min_{v \in \mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}_{\mathsf{x}}^{\mathsf{q}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\} \geq x\Big]$$

is continuous at $x = 0$ for any $t \in \mathbb{R}$. We utilize that fact in our proof of (2) and note in passing that most of the existing literature on monotone function estimators seems to implicitly utilize a similar continuity result.

A.3. Bootstrap inconsistency.
Employing the exchangeable bootstrap setup introduced in Section 4.2, we show the inconsistency of the plug-in bootstrap-based distributional approximation. More precisely, we consider the "naïve" (plug-in) bootstrap estimator

$$\widehat{\theta}_n^*(\mathsf{x}) = \partial_-\mathsf{GCM}_{[0,\widehat{u}_n^*]}(\widehat{\Gamma}_n^* \circ \widehat{\Phi}_n^{*-}) \circ \widehat{\Phi}_n^*(\mathsf{x})$$

with

$$\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} W_{i,n}\widehat{\gamma}_n(x; \mathbf{Z}_i), \qquad \widehat{\Phi}_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} W_{i,n}\widehat{\phi}_n(x; \mathbf{Z}_i),$$

and establish the following negative result.

THEOREM A.1. *Suppose Assumptions A, (B5)–(B7), and E hold. Then*

$$\sup_{t \in \mathbb{R}}|\mathbb{P}_n^*[\widehat{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x}) \leq t] - \mathbb{P}[\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x}) \leq t]| \neq o_{\mathbb{P}}(1).$$

Theorem A.1 accommodates exchangeable bootstrap schemes and a wide class of generalized Grenander-type estimators and contains as a special case the the well-known fact that the nonparametric bootstrap approximation to the distribution of the classical Grenander estimator is inconsistent (e.g., Kosorok (2008), Sen, Banerjee and Woodroofe (2010)).

A.4. Omitted formulas.
The objects $\widehat{V}_{\mathsf{x},n}^{\mathsf{q}}$, $M_{\mathsf{x},n}^{\mathsf{q}}$, $\widehat{L}_{\mathsf{x},n}^{\mathsf{q}}$, and $\widehat{Z}_{\mathsf{x},n}^{\mathsf{q}}$ appearing in Section 3.1 are defined as follows:

$$\widehat{V}_{\mathsf{x},n}^{\mathsf{q}} = \{a_n(x - \mathsf{x}) : x \in \widehat{\Phi}_n^-([0, \widehat{u}_n])\},$$

$$M_{\mathsf{x},n}^{\mathsf{q}}(v) = \sqrt{na_n}[\Gamma_0(\mathsf{x} + va_n^{-1}) - \Gamma_0(\mathsf{x})] - \theta_0(\mathsf{x})\sqrt{na_n}[\Phi_0(\mathsf{x} + va_n^{-1}) - \Phi_0(\mathsf{x})],$$

and

$$\widehat{L}_{\mathsf{x},n}^{\mathsf{q}}(v) = a_n[\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x})], \qquad \widehat{Z}_{\mathsf{x},n}^{\mathsf{q}} = a_n[(\widehat{\Phi}_n^- \circ \widehat{\Phi}_n)(\mathsf{x}) - \mathsf{x}].$$

Similarly, the bootstrap analogs of $\widehat{V}_{\mathsf{x},n}^{\mathsf{q}}$, $M_{\mathsf{x},n}^{\mathsf{q}}$, $\widehat{L}_{\mathsf{x},n}^{\mathsf{q}}$, and $\widehat{Z}_{\mathsf{x},n}^{\mathsf{q}}$ are given by

$$\widehat{V}_{\mathsf{x},n}^{\mathsf{q},*} = \{a_n(x - \mathsf{x}) : x \in \widehat{\Phi}_n^{*-}([0, \widehat{u}_n^*])\},$$

$$\widehat{M}_{\mathsf{x},n}^{\mathsf{q}}(v) = \sqrt{na_n}[\widehat{\Gamma}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x})] - \widehat{\theta}_n(\mathsf{x})\sqrt{na_n}[\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x})],$$

and

$$\widehat{L}_{\mathsf{x},n}^{\mathsf{q},*}(v) = a_n[\widehat{\Phi}_n^*(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n^*(\mathsf{x})], \qquad \widehat{Z}_{\mathsf{x},n}^{\mathsf{q},*} = a_n[(\widehat{\Phi}_n^{*-} \circ \widehat{\Phi}_n^*)(\mathsf{x}) - \mathsf{x}],$$

respectively.

SUPPLEMENTARY MATERIAL

Supplement to "Bootstrap-assisted infernce for generalized Grenander-type estimators" (DOI: 10.1214/24-AOS2402SUPP; .pdf). The supplementary material contains proofs of Theorems 1 and A.1, Corollaries 1–4, and Lemmas 1–3 and A.1–A.2. It discusses additional examples that our main results cover.

REFERENCES

BRUNK, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Stat.* **29** 437–454. MR0132632 https://doi.org/10.1214/aoms/1177706621

CATTANEO, M. D., JANSSON, M. and NAGASAWA, K. (2020). Bootstrap-based inference for cube root asymptotics. *Econometrica* **88** 2203–2219. MR4162697 https://doi.org/10.3982/ecta17950

CATTANEO, M. D., JANSSON, M. and NAGASAWA, K. (2024). Supplement to "Bootstrap-Assisted Inference for Generalized Grenander-type Estimators." https://doi.org/10.1214/24-AOS2402SUPP

CAVALIERE, G. and GEORGIEV, I. (2020). Inference under random limit bootstrap measures. *Econometrica* **88** 2547–2574. MR4192549 https://doi.org/10.3982/ECTA16557

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382 https://doi.org/10.1007/BF02868560

COX, G. (2022). A generalized argmax theorem with applications. Preprint. Available at arXiv:2209.08793.

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (*Berkeley, Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052

GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators, Algorithms and Asymptotics. Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893

GROENEBOOM, P. and JONGBLOED, G. (2018). Some developments in the theory of shape constrained inference. *Statist. Sci.* **33** 473–492. MR3881204 https://doi.org/10.1214/18-STS657

HAN, Q. and KATO, K. (2022). Berry-Esseen bounds for Chernoff-type nonstandard asymptotics in isotonic regression. *Ann. Appl. Probab.* **32** 1459–1498. MR4414710 https://doi.org/10.1214/21-aap1716

HONG, H. and LI, J. (2020). The numerical bootstrap. *Ann. Statist.* **48** 397–412. MR4065167 https://doi.org/10.1214/19-AOS1812

HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat.* (*IMS*) *Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212 https://doi.org/10.1214/193940307000000202

KUCHIBHOTLA, A. K., BALAKRISHNAN, S. and WASSERMAN, L.. (2024). The HulC: Confidence regions from convex hulls. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*

LEE, S. M. S. and YANG, P. (2020). Bootstrap confidence regions based on M-estimators under nonstandard conditions. *Ann. Statist.* **48** 274–299. MR4065162 https://doi.org/10.1214/18-AOS1803

MALLICK, S., SARKAR, S. and KUCHIBHOTLA, A. K. (2024). New asymptotic limit theory and inference for monotone regression. Preprint. Available at arXiv:2310.20058.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann*. *Statist*. **22** 2031–2050. MR1329181 https://doi.org/10.1214/aos/1176325770

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann*. *Statist*. **38** 1953–1977. MR2676880 https://doi.org/10.1214/09-AOS777

SEO, M. H. and OTSU, T. (2018). Local M-estimation with discontinuous criterion for dependent and limited observations. *Ann*. *Statist*. **46** 344–369. MR3766955 https://doi.org/10.1214/17-AOS1552

VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. *Springer Series in Statistics*. Springer, New York. MR1958123 https://doi.org/10.1007/978-0-387-21700-0

VAN DER VAART, A. and VAN DER LAAN, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int*. *J*. *Biostat*. **2** Art. 9, 42 pp. MR2306498 https://doi.org/10.2202/1557-4679.1014

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

WESTLING, T. and CARONE, M. (2020). A unified study of nonparametric inference for monotone functions. *Ann*. *Statist*. **48** 1001–1024. MR4102685 https://doi.org/10.1214/19-AOS1835

WESTLING, T., GILBERT, P. and CARONE, M. (2020). Causal isotonic regression. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **82** 719–747. MR4112782 https://doi.org/10.1111/rssb.12372