Bootstrap-Assisted Inference for Generalized Grenander-type Estimators*

Matias D. Cattaneo[†] Michael Jansson[‡] Kenichi Nagasawa[§]

December 27, 2022

Abstract

Westling and Carone (2020) proposed a framework for studying the large sample distributional properties of generalized Grenander-type estimators, a versatile class of nonparametric estimators of monotone function. The limiting distribution of those estimators is representable as the left derivative of the greatest convex minorant of a Gaussian process whose covariance kernel can be complicated and 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, which relies on a carefully crafted, yet automatic, transformation of the estimator. 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.

Keywords: monotone estimation, bootstrapping, robust inference.

*We are grateful to Boris Hanin, Jason Klusowski, Whitney Newey, Ted Westling, and seminar participants at various institutions for their comments. Cattaneo gratefully acknowledges financial support from the National Science Foundation through grant SES-1947805, and Jansson gratefully acknowledges financial support from the National Science Foundation through grant SES-1947662.

[†]Department of Operations Research and Financial Engineering, Princeton University.

[‡]Department of Economics, University of California at Berkeley.

[§]Department of Economics, University of Warwick.

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 so-called generalized Grenander-type estimators. Estimation problems covered by Westling and Carone's (2020) general theory contains many practically relevant examples such as monotone density, regression and hazard estimation, possibly with censoring and/or covariate adjustments.

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 covariance kernel and mean 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), the m-out-of-n bootstrap (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).

We complement existing methods by introducing a novel bootstrap-assisted inference approach that restores validity of bootstrap methods by reshaping the generalized Grenander-type estimator. Our approach is motivated by a constructive interpretation of the source of the failure of the nonparametric bootstrap. 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 non-standard 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 so-called *switch relation* (Groeneboom, 1985) to re-express the cumulative distribution function (cdf) of the centered and scaled monotone estimator in terms of a probability statement about the maximizer of a certain stochastic process, whose large sample properties in turn can 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 distribution of the maximizer of 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 the Generalized Switching Lemma of Westling and Carone (2020, Supplement) is incorrect as stated and then propose a modification. Relative to Westling and Carone's (2020, Supplement) formulation, our modification of the Generalized Switching Lemma includes one additional assumption, but removes some other unnecessary conditions. The additional assumption is satisfied in all examples considered by Westling and Carone (2020), as well as some new examples we considered in this paper. Second, we present a lemma establishing continuity of the cdf of the maximizer of a Gaussian process, a result which in turn can 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 Appendix A.

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. Appendix A reports the two technical results alluded to in the previous paragraph. All proofs and other technical derivations, as well as regularity conditions for the specific examples, are given in the supplemental appendix.

1.1 Notation and Definitions

For a function f defined on an interval $I \subseteq \mathbb{R}$, $\mathrm{GCM}_I(f)$ denotes its greatest convex minorant (on I) and f^- denotes its generalized inverse; that is, $f^-(x) = \inf\{u \in I : f(u) \geq x\}$. Assuming the relevant derivatives exist, ∂^q denotes the q-the partial derivative (operator) and ∂_- denotes the left derivative (operator). In addition, $f \circ g$ denotes the composition of f and g; that is, $(f \circ g)(x) = f(g(x))$.

Limits are taken as $n \to \infty$, unless otherwise stated. For 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 non-random. $a_n = o(b_n)$ denotes $a_n/b_n \to 0$, and $a_n = o_{\mathbb{P}}(b_n)$ denotes $a_n/b_n \to_{\mathbb{P}} 0$, where $\to_{\mathbb{P}}$ is convergence in probability. We use $\rightsquigarrow$ to denote weak convergence, where, for a stochastic process indexed by $\mathbb{R}$ or some subset thereof, convergence is in the topology of uniform convergence on compacta. When analyzing the bootstrap, $\mathbb{P}_n^*$ denotes the probability measure under the bootstrap sampling 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 $I \subseteq \mathbb{R}$, $\theta_0 : I \to \mathbb{R}$ is a non-decreasing function and x is an interior point of I. Assuming it is well defined, the function $\Theta_0 : I \to \mathbb{R}$ given by

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

is convex and therefore enojjys the property that if θ_0 is left continuous at x, then

$$\theta_0(\mathsf{x}) = \partial_- \mathsf{GCM}_I(\Theta_0)(\mathsf{x}). \tag{1}$$

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 estimator of a non-decreasing density.

Example 1 (Monotone Density Estimation). *Suppose $X_1, \dots, X_n$ are i.i.d. copies of a continuously distributed random variable X whose Lebesgue density f_0 is non-decreasing on its support $I = [0, u_0]$. 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 $\widehat{u}_n = \max(\max_{1 \le i \le n} X_i, \mathsf{x})$ and where $\widehat{F}_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}(X_i \le x)$ is the empirical cdf. Section 5.1 presents novel bootstrap-based inference methods for monotone density estimators possibly 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, for some $u_0 > 0$, $\Phi_0 : I \to [0, u_0]$ is a non-decreasing and right continuous function. Defining

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

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

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

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); that is, an estimator of Generalized Grenander-type is of the form

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

Of course, Grenander-type estimators are of generalized Grenander-type (obtained by setting Φ_0 and $\widehat{\Phi}_n$ equal to the identity mapping) whenever $I = [0, u_0]$, 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 $(Y_1, X_1), \dots, (Y_n, X_n)$ 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 non-decreasing. 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 $\widehat{\Gamma}_n(x) = n^{-1} \sum_{i=1}^n Y_i \mathbb{1}(X_i \le x)$ and $\widehat{F}_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}(X_i \le x)$. Section 5.2 presents novel bootstrap-based inference methods for monotone regression estimators possibly allowing for covariate adjustment (e.g., Westling, Gilbert and Carone, 2020).

Under regularity conditions, the rate of convergence of the generalized Grenander-type estimator

$\widehat{\theta}_n(\mathsf{x})$ is governed by the flatness of θ_0 around x, as measured by the characteristic exponent

$$\mathfrak{q} = \min\{j \in \mathbb{N} : \partial^j \theta_0(\mathsf{x}) \neq 0\}. \tag{2}$$

(When θ_0 is non-decreasing and suitably smooth, $\mathfrak{q}$ is necessarily an odd integer.) To be specific, Westling and Carone (2020, Theorem 3) gave conditions under which

$$r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})) \rightsquigarrow \frac{1}{\partial \Phi_0(\mathsf{x})} \partial_- \mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}})(0), \qquad r_n = n^{\mathfrak{q}/(1+2\mathfrak{q})}, \tag{3}$$

where $\mathcal{G}_{\mathsf{x}}$ is a zero mean Gaussian process and $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$ is a monomial (of order $\mathfrak{q}+1$) given by

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \frac{\partial^{\mathfrak{q}} \theta_0(\mathsf{x}) \partial \Phi_0(\mathsf{x})}{(\mathfrak{q}+1)!} v^{\mathfrak{q}+1}. \tag{4}$$

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 $\mathfrak{q} = 1$ and if $\mathcal{G}_{\mathsf{x}}$ is a two-sided Brownian motion, 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) whenever $\mathcal{G}_{\mathsf{x}}$ is a two-sided Brownian motion, but for more complicated $\mathcal{G}_{\mathsf{x}}$ the asymptotic distribution of $r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x}))$ need not belong to scale family of distributions.

The following assumptions guarantee validity of the representation (1) and ensures existence of the the right hand side of (4).

Assumption A. *For some $\delta > 0$, some $\mathfrak{s} \geq 1$, and some $\mathfrak{q} \in \mathbb{N}$, the following are satisfied:*

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

(A2) *$\theta_0 : I \to \mathbb{R}$ is non-decreasing. In addition, θ_0 is $\lfloor \mathfrak{s} \rfloor$ times continuously differentiable on I_{x}^{δ} with*

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

Also, $\mathfrak{q} \leq \lfloor \mathfrak{s} \rfloor$, where $\mathfrak{q}$ is the characteristic exponent defined in (2).

(A3) *$\Phi_0 : I \to [0, u_0]$ is non-decreasing and right continuous. In addition, Φ_0 is $\lfloor \mathfrak{s} \rfloor - \mathfrak{q} + 1$ 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

Letting $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ denote a bootstrap analog of $(\widehat{\Gamma}_n, \widehat{\Phi}_n, \widehat{u}_n)$, 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 (e.g. Kosorok, 2008; Sen et al., 2010), the following bootstrap analog of (3) does not (necessarily) hold when $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ is obtained by means of the nonparametric bootstrap:

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

In other words, the bootstrap is inconsistent in general. It turns out, however, that under plausible conditions on $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ a valid bootstrap-based distributional can be obtained by employing

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

where, for some judiciously chosen $\tilde{M}_{\mathsf{x},n}$, $\tilde{\Gamma}_n^*$ is given by

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

(As the notation suggests, a suitably rescaled $\tilde{M}_{\mathsf{x},n}$ can be interpreted as an estimator of $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$.)

3.1 Heuristics

To explain the source of the bootstrap failure and motivate the functional form of $\tilde{\Gamma}_n^*$, it is useful to begin by sketching the derivation of (3). Let $a_n = n^{1/(1+2\mathfrak{q})}$ and define

$$\begin{aligned} \widehat{G}_{\mathsf{x},n}^{\mathfrak{q}}(v) = &\sqrt{na_n}[\widehat{\Gamma}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x} + va_n^{-1}) + \Gamma_0(\mathsf{x})] \\ &- \theta_0(\mathsf{x})\sqrt{na_n}[\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x} + va_n^{-1}) + \Phi_0(\mathsf{x})], \end{aligned}$$

$$M_{\mathsf{x},n}^{\mathfrak{q}}(v) = \sqrt{na_n}[\Gamma_0(\mathsf{x} + va_n^{-1}) - \Gamma_0(\mathsf{x})] - \theta_0(\mathsf{x})\sqrt{na_n}[\Phi_0(\mathsf{x} + va_n^{-1}) - \Phi_0(\mathsf{x})],$$
$$\widehat{L}_{\mathsf{x},n}^{\mathfrak{q}}(v) = a_n[\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x})],$$

and

$$\widehat{Z}_{\mathsf{x},n}^{\mathfrak{q}} = a_n[\widehat{\Phi}_n^-(\widehat{\Phi}_n(\mathsf{x})) - \mathsf{x}].$$

For any $t \in \mathbb{R}$, assuming validity of the so-called switch relation, we have

$$\mathbb{P}\left[r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})) \le t\right] = \mathbb{P}\left[\arg\min_{v \in \widehat{V}_{\mathsf{x},n}^{\mathfrak{q}}} \{\widehat{G}_{\mathsf{x},n}^{\mathfrak{q}}(v) + M_{\mathsf{x},n}^{\mathfrak{q}}(v) - t\widehat{L}_{\mathsf{x},n}^{\mathfrak{q}}(v)\} - \widehat{Z}_{\mathsf{x},n}^{\mathfrak{q}} \le 0\right],$$

where $\widehat{V}_{\mathsf{x},n}^{\mathfrak{q}} = \{a_n(x - \mathsf{x}) : x \in I \cap \widehat{\Phi}_n^{-}([0, \widehat{u}_n])\}$. Assuming moreover that

$$(\widehat{G}_{\mathsf{x},n}^{\mathfrak{q}}, \widehat{L}_{\mathsf{x},n}^{\mathfrak{q}}, \widehat{Z}_{\mathsf{x},n}^{\mathfrak{q}}) \rightsquigarrow (\mathcal{G}_{\mathsf{x}}, \mathcal{L}_{\mathsf{x}}, 0), \qquad \mathcal{L}_{\mathsf{x}}(v) = v\partial\Phi_0(\mathsf{x}), \tag{5}$$

and

$$M_{\mathsf{x},n}^{\mathfrak{q}} \rightsquigarrow \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}, \tag{6}$$

it therefore stands to reason that under mild additional conditions we have

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left[r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})) \le t\right] &= \mathbb{P}\left[\arg\min_{v\in\mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\} \le 0\right] \\ &= \mathbb{P}\left[\frac{1}{\partial\Phi_0(\mathsf{x})}\partial_{-}\mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}})(0) \le t\right],\end{aligned}$$

the first equality following from the argmax continuous mapping theorem and the second equality being obtained by another application of the switch relation.

Next, let the natural bootstrap analogs of $\widehat{G}_{\mathsf{x},n}^{\mathfrak{q}}$, $M_{\mathsf{x},n}^{\mathfrak{q}}$, $\widehat{L}_{\mathsf{x},n}^{\mathfrak{q}}$, and $\widehat{Z}_{\mathsf{x},n}^{\mathfrak{q}}$ be given by

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

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

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

and

$$\widehat{Z}_{\mathsf{x},n\mathsf{x}}^{\mathfrak{q},*} = a_n[(\widehat{\Phi}_n^*)^{-}(\widehat{\Phi}_n^*(\mathsf{x})) - \mathsf{x}],$$

respectively. For every $t \in \mathbb{R}$, it follows from the switch relation that these objects satisfy

$$\mathbb{P}_n^*\left[r_n(\widehat{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})) \le t\right] = \mathbb{P}_n^*\left[\arg\min_{v\in\widehat{V}_{\mathsf{x},n}^{\mathfrak{q},*}}\{\widehat{G}_{\mathsf{x},n}^{\mathfrak{q},*}(v) + \widehat{M}_{\mathsf{x},n}^{\mathfrak{q}}(v) - t\widehat{L}_{\mathsf{x},n}^{\mathfrak{q},*}(v)\} - \widehat{Z}_{\mathsf{x},n}^{\mathfrak{q},*} \le 0\right],$$

where $\widehat{V}_{\mathsf{x},n}^{\mathfrak{q},*} = \{a_n(x - \mathsf{x}) : x \in I \cap (\widehat{\Phi}_n^*)^{-}([0, \widehat{u}_n^*])\}$. When (5) holds, then the following bootstrap counterpart thereof can also be expected to hold provided that $(\widehat{\Gamma}_n^*, \widehat{\Phi}_n^*, \widehat{u}_n^*)$ is chosen appropriately:

$$(\widehat{G}_{\mathsf{x},n}^{\mathfrak{q},*}, \widehat{L}_{\mathsf{x},n}^{\mathfrak{q},*}, \widehat{Z}_{\mathsf{x},n}^{\mathfrak{q},*}) \rightsquigarrow_{\mathbb{P}} (\mathcal{G}_{\mathsf{x}}, \mathcal{L}_{\mathsf{x}}, 0). \tag{7}$$

On the other hand, the process $\widehat{M}_{\mathsf{x},n}^{\mathfrak{q}}$ typically does not inherit the smoothness of $M_{\mathsf{x},n}^{\mathfrak{q}}$ and the bootstrap counterpart of (6) typically fails (e.g., Sen et al., 2010), implying in turn that the bootstrap is inconsistent.

By construction, the estimator $\tilde{\theta}_n^*(\mathsf{x})$ is similar to $\widehat{\theta}_n^*(\mathsf{x})$ insofar as it satisfies the following switch relation: For every $t \in \mathbb{R}$,

$$\mathbb{P}_n^*\left[r_n(\tilde{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})) \leq t\right] = \mathbb{P}_n^*\left[\arg\min_{v \in \widehat{V}_{\mathsf{x},n}^{\mathfrak{q},*}} \{\widehat{G}_{\mathsf{x},n}^{\mathfrak{q},*}(v) + \tilde{M}_{\mathsf{x},n}^{\mathfrak{q}}(v) - t\widehat{L}_{\mathsf{x},n}^{\mathfrak{q},*}(v)\} - \widehat{Z}_{\mathsf{x},n}^{\mathfrak{q},*} \leq 0\right],$$

where

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

As a consequence, if (7) holds and if $\tilde{M}_{\mathsf{x},n}^{\mathfrak{q}} \rightsquigarrow_{\mathbb{P}} \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$, then it stands to reason that under mild additional conditions $\tilde{\theta}_n^*(\mathsf{x})$ satisfies the following bootstrap counterpart of (3):

$$r_n(\tilde{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})) \rightsquigarrow_{\mathbb{P}} \frac{1}{\partial\Phi_0(\mathsf{x})}\partial_-\mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}})(0). \tag{8}$$

3.2 Main result

Our heuristic derivation of (3) can be made rigorous by providing conditions under which four properties hold. First, the switch relation(s) must be valid. Second, the convergence properties (5) and (6) must hold. Third, to use (5) and (6) to obtain the result

$$\arg\min_{v \in \widehat{V}_{\mathsf{x},n}^{\mathfrak{q}}} \{\widehat{G}_{\mathsf{x},n}^{\mathfrak{q}}(v) + M_{\mathsf{x},n}^{\mathfrak{q}}(v) - t\widehat{L}_{\mathsf{x},n}^{\mathfrak{q}}(v)\} - \widehat{Z}_{\mathsf{x},n}^{\mathfrak{q}} \rightsquigarrow \arg\min_{v \in \mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\}$$

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

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left[\arg\min_{v \in \widehat{V}_{\mathsf{x},n}^{\mathfrak{q}}} \{\widehat{G}_{\mathsf{x},n}^{\mathfrak{q}}(v) + M_{\mathsf{x},n}^{\mathfrak{q}}(v) - t\widehat{L}_{\mathsf{x},n}^{\mathfrak{q}}(v)\} - \widehat{Z}_{\mathsf{x},n}^{\mathfrak{q}} \leq 0\right] \\
&= \mathbb{P}\left[\arg\min_{v \in \mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\} \leq 0\right],
\end{aligned}$$

the cdf of $\arg\min_{v\in\mathbb{R}}\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) - t\mathcal{L}_{\mathsf{x}}(v)\}$ 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. For concreteness, we base our formulations on van der Vaart and Wellner (1996). The first and fourth properties, on the other hand, seem more difficult to verify. Regarding the first property, although various statements of versions of the switch relation can be found in the literature, we are unaware of a version of the result which can handle reasonably general examples of $\widehat{\Phi}_n$ and $\widehat{\Phi}_n^*$. In particular, it turns out that the version of the switch relation employed by Westling and Carone (2020) is incorrect as stated, but thankfully it turns out that a seemingly mild extra condition (namely, closedness of the range of $\widehat{\Phi}_n$ and $\widehat{\Phi}_n^*$) is enough to establish/restore validity of the switch relation stated in Lemma 1 in the supplemental appendix of Westling and Carone (2020). For details, see Lemma A.1 in the appendix. In the special case where $\mathfrak{q} = 1$ and $\mathcal{G}_{\mathsf{x}}$ is a two-sided Brownian motion, 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 $\mathfrak{q} \neq 1$ and/or $\mathcal{G}_{\mathsf{x}}$ is not a two-sided Brownian motion, but fortunately it turns out that the continuity property of interest admits simple sufficient conditions (namely, **(B4)** in Assumption B). For details, see Lemma A.2 in the appendix.

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

Assumption B. *For the same* $\mathfrak{q}$ *as in Assumption A, the following are satisfied:*

(B1) $\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} \rightsquigarrow \mathcal{G}_{\mathsf{x}}$ *and* $\widehat{G}^{\mathfrak{q},*}_{\mathsf{x},n} \rightsquigarrow_{\mathbb{P}} \mathcal{G}_{\mathsf{x}}$, *where* $\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n}$ *and* $\widehat{G}^{\mathfrak{q},*}_{\mathsf{x},n}$ *are defined as in Section 3.1 and where* $\mathcal{G}_{\mathsf{x}}$ *is a zero mean Gaussian process.*

(B2) $\sup_{x\in I} |\widehat{\Gamma}_n(x) - \Gamma_0(x)| = o_{\mathbb{P}}(1)$ *and* $\sup_{x\in I} |\widehat{\Gamma}^*_n(x) - \widehat{\Gamma}_n(x)| = o_{\mathbb{P}}(1)$.

(B3) $\widehat{\Phi}_n$ *and* $\widehat{\Phi}^*_n$ *are non-decreasing, right continuous, and have closed range.*

In addition, $\sup_{x\in I} |\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1)$ *and* $\sup_{x\in I} |\widehat{\Phi}^*_n(x) - \widehat{\Phi}_n(x)| = o_{\mathbb{P}}(1)$. *Also, for every* $K > 0$,

$$a_n \sup_{|v|\leq K} |\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x} + va_n^{-1}) + \Phi_0(\mathsf{x})| = o_{\mathbb{P}}(1)$$

and

$$a_n \sup_{|v|\leq K} |\widehat{\Phi}^*_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}^*_n(\mathsf{x}) - \widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) + \widehat{\Phi}_n(\mathsf{x})| = o_{\mathbb{P}}(1).$$

(B4) *For every* $v, v' \in \mathbb{R}$, *the covariance kernel* $\mathcal{C}_{\mathsf{x}}$ *of* $\mathcal{G}_{\mathsf{x}}$ *satisfies*

$$\mathcal{C}_{\mathsf{x}}(v + v', v + v') - \mathcal{C}_{\mathsf{x}}(v + v', v') - \mathcal{C}_{\mathsf{x}}(v', v + v') + \mathcal{C}_{\mathsf{x}}(v', v') = \mathcal{C}_{\mathsf{x}}(v, v)$$

and

$$\mathcal{C}_{\mathsf{x}}(v\tau, v'\tau) = \mathcal{C}_{\mathsf{x}}(v, v')\tau \qquad \text{for every } \tau \geq 0.$$

In addition, $\mathcal{C}_{\mathsf{x}}(1, 1) > 0$ *and* $\lim_{\delta\downarrow 0} \mathcal{C}_{\mathsf{x}}(1, \delta)/\sqrt{\delta} = 0$.

(B5) $|\widehat{u}_n - u_0| = o_{\mathbb{P}}(1)$ *and* $|\widehat{u}^*_n - \widehat{u}_n| = o_{\mathbb{P}}(1)$.

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 (3) and (7). In addition, (8) can be shown to hold if $\tilde{M}_{\mathsf{x},n}$ satisfies the following

Assumption C. $\tilde{M}^{\mathfrak{q}}_{\mathsf{x},n} \rightsquigarrow_{\mathbb{P}} \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$ *and, for every* $K > 0$,

$$\liminf_{\delta\downarrow 0} \liminf_{n\to\infty} \mathbb{P}\left[\inf_{|v|>K^{-1}} \tilde{M}_{\mathsf{x},n}(v) \geq \delta\right] = 1.$$

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

Theorem 1. *Suppose Assumptions A, B, and C are satisfied. Then (3) and (8) hold. In addition,*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}_n^*\left[\tilde{\theta}_n^*(\mathsf{x})-\widehat{\theta}_n(\mathsf{x})\leq t\right]-\mathbb{P}\left[\widehat{\theta}_n(\mathsf{x})-\theta_0(\mathsf{x})\leq t\right]\right|=o_{\mathbb{P}}(1). \tag{9}$$

In an attempt to emphasize the rate-adaptive nature of the consistency property enjoyed by the bootstrap-based distributional approximation based on $\tilde{\theta}_n^*(\mathsf{x})$, the formulation (9) deliberately omits the rate term r_n present in (3) and (8). Theorem 1 has immediate implications for inference. For instance, it follows from (3) and (9) that for any $\alpha\in(0,1)$, we have

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

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

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

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 the triple $(\widehat{\Gamma}_n,\widehat{\Phi}_n,\widehat{u}_n)$ satisfies the non-bootstrap parts of Assumption B. Then, in order to compute the estimator $\tilde{\theta}_n^*(\mathsf{x})$ upon which our proposed bootstrap-based distributional approximation is based, two implementational issues must be addressed, namely the choice/construction of $\tilde{M}_{\mathsf{x},n}$ and $(\widehat{\Gamma}_n^*,\widehat{\Phi}_n^*,\widehat{u}_n^*)$, respectively. Section 4.1 demonstrates the plausibility of Assumption C by exhibiting estimators $\tilde{M}_{\mathsf{x},n}$ satisfying it under Assumptions A and D, the latter being a high-level condition that typicaly holds whenever the non-bootstrap 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 $\tilde{M}_{\mathsf{x},n}$ satisfying Assumption C can be constructed depends on whether $\mathfrak{q}$ is known. To facilitate the discussion, for $j=1,\ldots,\lfloor\mathfrak{s}\rfloor$, let

$$\mathcal{D}_j(\mathsf{x})=\frac{\partial^{j+1}\Upsilon_0(\mathsf{x})}{(j+1)!},\qquad\Upsilon_0(x)=\Gamma_0(x)-\theta_0(\mathsf{x})\Phi_0(x).$$

As defined, Υ_0 satisfies $\partial\Upsilon_0(\mathsf{x})=\ldots=\partial^{\mathfrak{q}}\Upsilon_0(\mathsf{x})=0$ and

$$\partial^{j+1}\Upsilon_0(\mathsf{x})=\sum_{k=\mathfrak{q}}^{j}\binom{j}{k}\partial^k\theta_0(\mathsf{x})\partial^{j+1-k}\Phi_0(\mathsf{x}),\qquad j=\mathfrak{q},\ldots,\lfloor\mathfrak{s}\rfloor\,,$$

implying in particular that $\partial^{\mathfrak{q}+1}\Upsilon_0(\mathsf{x}) = \partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(\mathsf{x})$ and therefore $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \mathcal{D}_{\mathfrak{q}}(\mathsf{x})v^{\mathfrak{q}+1}$.

First, consider the (simpler) case where $\mathfrak{q}$ is known. In this case, if

$$\tilde{\mathcal{D}}_{\mathfrak{q},n}(\mathsf{x}) \to_{\mathbb{P}} \mathcal{D}_{\mathfrak{q}}(\mathsf{x}), \tag{10}$$

then Assumption C holds when

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

Examples of estimators $\tilde{\mathcal{D}}_{\mathfrak{q},n}(\mathsf{x})$ satisfying the consistency requirement (10) will be given below.

Next, consider the somewhat more complicated case where $\mathfrak{q}$ is unknown, but assumed to satisfy $\mathfrak{q} \leq \bar{\mathfrak{q}}$ for some known integer $\bar{\mathfrak{q}} \geq 3$. Noting that $\mathcal{D}_j(\mathsf{x}) = 0$ for $j < \mathfrak{q}$, $\mathcal{D}_{\mathfrak{q}}(\mathsf{x}) > 0$, and that $\mathfrak{q}$ is necessarily an odd integer, $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$ can be written as

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \sum_{\ell=1}^{\lfloor(\bar{\mathfrak{q}}+1)/2\rfloor} \mathbb{1}(2\ell \leq \mathfrak{q}) \max(\mathcal{D}_{2\ell-1}(\mathsf{x}), 0)v^{2\ell}.$$

Dropping the indicator function term from each summand, we obtain

$$\bar{\mathcal{M}}_{\mathsf{x}}(v) = \sum_{\ell=1}^{\lfloor(\bar{\mathfrak{q}}+1)/2\rfloor} \max(\mathcal{D}_{2\ell-1}(\mathsf{x}), 0)v^{2\ell}.$$

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

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

Moreover, the following ("global") positivity property automatically holds:

$$\inf_{|v|>K^{-1}} \bar{\mathcal{M}}_{\mathsf{x}}(v) > 0, \qquad \text{for every } K > 0.$$

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

$$a_n^{\mathfrak{q}-(2\ell-1)}(\tilde{\mathcal{D}}_{2\ell-1,n}(\mathsf{x}) - \mathcal{D}_{2\ell-1}(\mathsf{x})) = o_{\mathbb{P}}(1), \qquad \ell = 1, \ldots, \lfloor(\bar{\mathfrak{q}}+1)/2\rfloor, \tag{11}$$

then Assumption C is satisfied by

$$\tilde{M}_{\mathsf{x},n}(v) = \sum_{\ell=1}^{\lfloor(\bar{\mathfrak{q}}+1)/2\rfloor} \max(\tilde{\mathcal{D}}_{2\ell-1,n}(\mathsf{x}), 0)v^{2\ell}.$$

For "small" ℓ (namely, for $2\ell - 1 < \mathfrak{q}$), the precision requirement (11) is stronger than consistency, but fortunately it turns out that the requirement can be met as long as $\mathfrak{s}$ is larger than $\bar{\mathfrak{q}}$.

Example-specific estimators $\tilde{\mathcal{D}}_{\mathfrak{q},n}(\mathsf{x})$ satisfying the consistency requirement (10) are sometimes readily available. For instance, in the case of the Grenander (1956) estimator (i.e., in Example 1), we have $\mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \partial^{\mathfrak{q}} f_0(\mathsf{x})/(\mathfrak{q}+1)!$, a consistent estimator of which can be based on any consistent estimator of $\partial^{\mathfrak{q}} f_0(\mathsf{x})$, 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) non-smooth estimator of Υ_0:

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

For $j = 1, \ldots, \mathfrak{q}$, the fact that $\partial \Upsilon_0(\mathsf{x}) = \ldots = \partial^{\mathfrak{q}} \Upsilon_0(\mathsf{x}) = 0$ implies that $\mathcal{D}_j(\mathsf{x})$ admits the "monomial approximation" representation

$$\mathcal{D}_j(\mathsf{x}) = \lim_{\epsilon \to 0} \left\{ \epsilon^{-(j+1)} [\Upsilon_0(\mathsf{x}+\epsilon) - \Upsilon_0(\mathsf{x})] \right\},$$

motivating the estimator

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

where $\epsilon_n > 0$ is a (small) tuning parameter. Similarly, for any $j = 1, \ldots, \lfloor \mathfrak{s} \rfloor$, the generic "forward difference" representation

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

motivates the estimator

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

Finally, to define an estimator with (possibly) superior bias properties, suppose $\mathfrak{s}$ admits a known integer $\underline{\mathfrak{s}}$ satisfying $\underline{\mathfrak{s}} \leq \mathfrak{s}$ and let the defining property of $\{\lambda^{\mathtt{BR}}_j(k) : 1 \leq k \leq \underline{\mathfrak{s}}+1\}$ be

$$\sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda^{\mathtt{BR}}_j(k) k^p = \mathbb{1}(p = j+1), \qquad p = 1, \ldots, \underline{\mathfrak{s}}+1.$$

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

$$\tilde{\mathcal{D}}^{\mathtt{BR}}_{j,n}(\mathsf{x}) = \epsilon_n^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda^{\mathtt{BR}}_j(k) [\widehat{\Upsilon}_n(\mathsf{x}+k\epsilon_n) - \widehat{\Upsilon}_n(\mathsf{x})]$$

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

$$\mathcal{D}_j(\mathsf{x}) \approx \epsilon^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)[\Upsilon_0(\mathsf{x}+k\epsilon)-\Upsilon_0(\mathsf{x})]$$

is of order $\epsilon^{\min(\underline{\mathfrak{s}}+1,\mathfrak{s})-j}$ when $\mathfrak{s} > j$. Relative to $\tilde{\mathcal{D}}_{j,n}^{\mathtt{MA}}(\mathsf{x})$ and $\tilde{\mathcal{D}}_{j,n}^{\mathtt{FD}}(\mathsf{x})$, this is a distinguishing feature of $\tilde{\mathcal{D}}_{j,n}^{\mathtt{BR}}(\mathsf{x})$ and as it turns out this feature will enable us to formulate sufficient conditions for (11).

For $\delta > 0$, let

$$\begin{aligned}\widehat{G}_{\mathsf{x},n}(v;\delta) &= \sqrt{n\delta^{-1}}[\widehat{\Gamma}_n(\mathsf{x}+v\delta)-\widehat{\Gamma}_n(\mathsf{x})-\Gamma_0(\mathsf{x}+v\delta)+\Gamma_0(\mathsf{x})] \\ &\quad -\theta_0(\mathsf{x})\sqrt{n\delta^{-1}}[\widehat{\Phi}_n(\mathsf{x}+v\delta)-\widehat{\Phi}_n(\mathsf{x})-\Phi_0(\mathsf{x}+v\delta)+\Phi_0(\mathsf{x})],\end{aligned}$$

and

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

Using this notation, the first part of **(B1)** and the first displayed condition of **(B3)** can be restated as

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

and

$$\sup_{|v|\leq K} |\widehat{R}_{\mathsf{x},n}(v;a_n^{-1})| = o_{\mathbb{P}}(1) \qquad \text{for every } K>0,$$

respectively. In the displayed results, one can typically replace $a_n^{-1} = n^{-1/(1+2\mathfrak{q})}$ by any $\delta_n > 0$ with $\delta_n = o(1)$ and $a_n^{-1}\delta_n^{-1} = O(1)$. As a consequence, validity of the following assumption usually follows as a by-product of the arguments used to justify **(B1)** and **(B3)**. (An illustration of this phenomenon is provided by Lemma 2 below.)

Assumption D. *For the same $\mathfrak{q}$ as in Assumption A and for every $\delta_n > 0$ with $\delta_n = o(1)$ and $a_n^{-1}\delta_n^{-1} = O(1)$,*

$$\widehat{G}_{\mathsf{x},n}(1;\delta_n) = O_{\mathbb{P}}(1) \qquad \textit{and} \qquad \widehat{R}_{\mathsf{x},n}(1;\delta_n) = o_{\mathbb{P}}(1).$$

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

Lemma 1. *Suppose Assumptions A and D are satisfied and that $r_n(\widehat{\theta}_n(\mathsf{x})-\theta_0(\mathsf{x})) = O_{\mathbb{P}}(1)$. If $\epsilon_n \to 0$ and if $a_n\epsilon_n \to \infty$, then*

$$\tilde{\mathcal{D}}_{\mathfrak{q},n}(\mathsf{x}) \to_{\mathbb{P}} \mathcal{D}_{\mathfrak{q}}(\mathsf{x}), \qquad \tilde{\mathcal{D}}_{\mathfrak{q},n} \in \{\tilde{\mathcal{D}}_{\mathfrak{q},n}^{\mathtt{MA}}, \tilde{\mathcal{D}}_{\mathfrak{q},n}^{\mathtt{FD}}, \tilde{\mathcal{D}}_{\mathfrak{q},n}^{\mathtt{BR}}\}$$

and

$$a_n^{\mathfrak{q}-j}(\tilde{\mathcal{D}}_{j,n}^{\mathtt{BR}}(\mathsf{x})-\mathcal{D}_j(\mathsf{x})) = O(a_n^{\mathfrak{q}-j}\epsilon_n^{\min(\underline{\mathfrak{s}}+1,\mathfrak{s})-j}) + o_{\mathbb{P}}(1), \qquad j=1,\ldots,\underline{\mathfrak{s}}.$$

In particular, if $3 \leq \overline{\mathfrak{q}} < \mathfrak{s}$, then (11) is satisfied by $\tilde{\mathcal{D}}_{2\ell-1,n} = \tilde{\mathcal{D}}_{2\ell-1,n}^{\mathtt{BR}}$ if $n\epsilon_n^{(1+2\overline{\mathfrak{q}})\min(\underline{\mathfrak{s}},\mathfrak{s}-1)/(\overline{\mathfrak{q}}-1)} \to 0$ and if $n\epsilon_n^{1+2\overline{\mathfrak{q}}} \to \infty$.

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

$$a_n^{\mathfrak{q}-j}(\tilde{\mathcal{D}}_{j,n}(\mathsf{x}) - \mathcal{D}_j(\mathsf{x})) = O(a_n^{\mathfrak{q}-j}\epsilon_n^{\mathfrak{q}-j}) + o_{\mathbb{P}}(1), \qquad j = 1, \ldots, \mathfrak{q}-1,$$

the $O(a_n^{\mathfrak{q}-j}\epsilon_n^{\mathfrak{q}-j})$ 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 non-negligible and the estimators $\tilde{\mathcal{D}}^{\texttt{MA}}$ and $\tilde{\mathcal{D}}^{\texttt{FD}}$ therefore do not satisfy (11).

Under additional assumptions (including $\mathfrak{s} \geq \underline{\mathfrak{s}} + 1$ and additional smoothness on Φ_0), $\tilde{\mathcal{D}}^{\texttt{BR}}$ 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

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

where the bias constant is

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

and the variance constant is

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

For details, see Section SA.5.4 in the supplemental appendix. Assuming $\mathsf{B}_j^{\texttt{BR}}(\mathsf{x}) \neq 0$, the approximate MSE is minimized by

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

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 $\mathsf{B}_j^{\texttt{BR}}(\mathsf{x})$ and $\mathsf{V}_j^{\texttt{BR}}(\mathsf{x})$, respectively.

Example 1 (Monotone Density Estimation, continued)**.** *In this example,*

$$\mathcal{D}_{\underline{\mathfrak{s}}+1}(\mathsf{x}) = \frac{1}{(\underline{\mathfrak{s}}+2)!}\partial^{\underline{\mathfrak{s}}+1}f_0(\mathsf{x}) \qquad \textit{and} \qquad \mathcal{C}_{\mathsf{x}}(k,l) = f_0(\mathsf{x})\min(k,l),$$

so a feasible version of $\epsilon_{j,n}^{\texttt{AMSE}}$ can be based on estimators of $\partial^{\underline{\mathfrak{s}}+1}f_0(\mathsf{x})$ and f_0. One option is to obtain consistent estimators by employing standard nonparametric techniques. Alternatively, a Silverman-style approach would obtain a feasible version of $\epsilon_{j,n}^{\texttt{BR}}(\mathsf{x})$ by working with a suitable reference distribution.

Example 2 (Monotone Regression Estimation, continued)**.** *In this example,*

$$\mathcal{D}_{\underline{\mathfrak{s}}+1}(\mathsf{x}) = \frac{1}{(\underline{\mathfrak{s}}+2)!}\sum_{k=\mathfrak{q}}^{\underline{\mathfrak{s}}+1}\binom{\underline{\mathfrak{s}}+1}{k}\partial^k\mu_0(\mathsf{x})\partial^{\underline{\mathfrak{s}}+1-k}f_0(\mathsf{x}) \qquad \text{and} \qquad \mathcal{C}_{\mathsf{x}}(k,l) = \sigma_0^2(\mathsf{x})f_0(\mathsf{x})\min(k,l),$$

where f_0 is the Lebesgue density of X and where $\sigma_0^2(x) = \mathbb{V}(Y|X=x)$. Again, one can obtain consistent estimators of the unknown components of $\mathcal{D}_{\underline{\mathfrak{s}}+1}(\mathsf{x})$ and $\mathcal{C}_{\mathsf{x}}(k,l)$ by employing standard nonparametric techniques or one can adopt a Silverman-style and obtain a feasible version of $\epsilon^{\mathtt{BR}}_{j,n}(\mathsf{x})$ by working with a suitable reference model.

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

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

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

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

respectively. Finally, $\widehat{u}_n = \max(\max_{1\le i\le n}\mathbf{Z}_i,\mathsf{x})$.

Letting $\mathbf{Z}^*_{1,n},\ldots,\mathbf{Z}^*_{n,n}$ denote a random sample from the empirical distribution of $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$, the nonparametric bootstrap analog $\widehat{u}_n$ is given by $\widehat{u}_n^* = \max(\max_{1\le i\le n}\mathbf{Z}^*_{i,n},\mathsf{x})$, while

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

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 part of **(B5)**. As for $\widehat{\Gamma}_n$ and $\widehat{\Phi}_n$, their nonparametric bootstrap analogs admit the weighted bootstrap representations*

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

where, conditionally on $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, $(W_{1,n}, \ldots, W_{n,n})$ *is multinomially distributed with probabilities* $(1/n, \ldots, 1/n)$ *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, the ability to satisfy the bootstrap part of **(B5)** by setting $\widehat{u}_n^* = \widehat{u}_n$ is of course generic and as a practical matter this choice seems as good as any. 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 non-bootstrap 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 $\Phi_0(x) = \mathbb{E}[\phi_0(x; \mathbf{Z})]$) for which the errors of the approximations

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

are suitably small. 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 weighted bootstrap-type pair

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

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

To give a precise statement, let

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

and define

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

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

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

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 where the supremum is over all discrete probability measure Q with $\|\bar{F}\|_{Q,2} > 0$.

Assumption E. *For the same $\mathfrak{q}$ as in Assumption A, the following are satisfied:*

(E1) $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ *are independent and identically distributed.*

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

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

(E3) $\sup_{x \in I} |\widehat{\Gamma}_n(x) - \Gamma_0(x)| = o_{\mathbb{P}}(1)$, $n^{-1}\sum_{i=1}^{n} \sup_{x \in I} |\widehat{\gamma}_n(x; \mathbf{Z}_i) - \gamma_0(x; \mathbf{Z}_i)|^2 = o_{\mathbb{P}}(1)$, *and, for every $K > 0$,*

$$\sqrt{na_n} \sup_{|v| \le K} \left|\widehat{\Gamma}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x} + va_n^{-1}) + \bar{\Gamma}_n(\mathsf{x})\right| = o_{\mathbb{P}}(1)$$

and

$$\frac{a_n}{n}\sum_{i=1}^{n} \sup_{|v| \le K} \left|\widehat{\gamma}_n(\mathsf{x} + va_n^{-1}; \mathbf{Z}_i) - \widehat{\gamma}_n(\mathsf{x}; \mathbf{Z}_i) - \gamma_0(\mathsf{x} + va_n^{-1}; \mathbf{Z}_i) + \gamma_0(\mathsf{x}; \mathbf{Z}_i)\right|^2 = o_{\mathbb{P}}(1).$$

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

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

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,

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

where $\bar{D}_\gamma^\delta$ is the minimal envelope of $\{\gamma_0(x; \cdot) - \gamma_0(\mathsf{x}; \cdot) : x \in I_{\mathsf{x}}^\delta\}$.

(E4) $\widehat{\Phi}_n, \widehat{\Phi}_n^*$ *are non-decreasing, right continuous, and have closed range.*

Also, $\sup_{x \in I} |\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1)$, $n^{-1}\sum_{i=1}^{n} \sup_{x \in I} |\widehat{\phi}_n(x; \mathbf{Z}_i) - \phi_0(x; \mathbf{Z}_i)|^2 = o_{\mathbb{P}}(1)$ *and, for every $K > 0$,*

$$\sqrt{na_n} \sup_{|v| \le K} \left|\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x} + va_n^{-1}) + \bar{\Phi}_n(\mathsf{x})\right| = o_{\mathbb{P}}(1)$$

and

$$\frac{a_n}{n}\sum_{i=1}^{n} \sup_{|v| \le K} \left|\widehat{\phi}_n(\mathsf{x} + va_n^{-1}; \mathbf{Z}_i) - \widehat{\phi}_n(\mathsf{x}; \mathbf{Z}_i) - \phi_0(\mathsf{x} + va_n^{-1}; \mathbf{Z}_i) + \phi_0(\mathsf{x}; \mathbf{Z}_i)\right|^2 = o_{\mathbb{P}}(1).$$

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

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

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,*

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

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

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

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

and, for all $v, v' \in \mathbb{R}$*, and for some* $\mathcal{C}_{\mathsf{x}}$*,*

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

Lemma 2. *Suppose Assumptions A and E are satisfied. Then* ***(B1)-(B3)*** *are satisfied. If also*

$$\sqrt{n\delta_n^{-1}}[\widehat{\Gamma}_n(\mathsf{x}+\delta_n) - \widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x}+\delta_n) + \bar{\Gamma}_n(\mathsf{x})] = O_{\mathbb{P}}(1)$$

for every $\delta_n > 0$ *with* $\delta_n = o(1)$ *and* $a_n^{-1}\delta_n^{-1} = O(1)$*, then Assumption D is satisfied.*

If Lemma 2 is used to verify **(B1)-(B3)**, then **(B4)** can usually be verified with minimal additional effort. In fact, the second displayed part of **(B4)** is implied by the second displayed part of **(E5)** and the first displayed part of **(B4)** is implied by the following locally uniform (with respect to x) strengthening of the second displayed part of **(E5)**:

$$\sup_{x \in I_{\mathsf{x}}^{\delta_n}} \left| \frac{\mathbb{E}[\psi_x(v\delta_n;\mathbf{Z})\psi_x(v'\delta_n;\mathbf{Z})]}{\delta_n} - \mathcal{C}_{\mathsf{x}}(v,v') \right| \to 0.$$

Moreover, it is usually not difficult to verify that $\mathcal{C}_{\mathsf{x}}$ satisfies the non-degeneracy condition $\mathcal{C}_{\mathsf{x}}(1,1) > 0$ and the (Hölder-type) continuity condition $\lim_{\delta \downarrow 0} \mathcal{C}_{\mathsf{x}}(1,\delta)/\sqrt{\delta} = 0$.

5 Examples

We apply our main results to two distinct sets of examples, both previously analyzed in Westling and Carone (2020) and Westling et al. (2020), and references therein. In the supplemental appendix, we also consider two other set of examples: Monotone Hazard Estimation (Huang and Wellner,

1995) and Monotone Distribution Estimation (van der Vaart and van der Laan, 2006). To conserve space, this section only offers an overview of our main results for each of the examples. Precise regularity conditions are stated in the supplemental appendix.

The examples considered rely on the same basic setup. Let $\mathbf{Z}_i = (Y_i, \check{X}_i, \Delta_i, \mathbf{A}_i)'$, $i = 1, 2, \ldots, n$, be an observed random sample with $\check{X}_i = \min\{X_i, C_i\}$, $\Delta_i = \mathbb{1}(X_i \leq C_i)$, $\mathbf{A}_i$ denoting additional covariates. If $\mathbb{P}[C_i \geq X_i] = 1$, then there is no (right) censoring and $\check{X}_i = X_i$. Assuming that $F_0(x) = \mathbb{P}[X_i \leq x]$ is absolutely continuous, $f_0(x)$ denotes the Lebesgue density of X. Letting $\mu_0(x) = \mathbb{E}[Y_i|X_i = x]$, the examples consider monotone estimation of $f_0(\mathsf{x})$ and $\mu_0(\mathsf{x})$, respectively, under various assumptions related to censoring and covariate-adjustment.

5.1 Monotone Density Estimation

As a first set of examples, consider the problem of estimating a monotone density of a continuously distributed random variable, possibly with censoring and covariate-adjustment. Let $\mathbf{Z}_i = (\check{X}_i, \Delta_i, \mathbf{A}_i)'$, $i = 1, 2, \ldots, n$, be an observed random sample with $\check{X}_i = \min\{X_i, C_i\}$, $\Delta_i = \mathbb{1}(X_i \leq C_i)$, and $\mathbf{A}_i$ denoting additional covariates. Assuming that f_0, the Lebesgue density of X, exists and is non-decreasing on $I = [0, u_0]$, the parameter of interest is $\theta_0(\mathsf{x}) = f_0(\mathsf{x})$ for some $\mathsf{x} \in (0, u_0)$.

Throughout, we set $\widehat{\Phi}_0(x) = \widehat{\Phi}_n(x) = x$ and, if u_0 is unknown, $\widehat{u}_n = \max(\max_{1\leq i\leq n} X_i, \mathsf{x})$. Similarly, we set $\widehat{\Phi}_n^*(x) = x$ and $\widehat{u}_n^* = \widehat{u}_n$. It remains to specify $\widehat{\Gamma}_n$ and $\widehat{\Gamma}_n^*$.

The canonical case of no censoring (i.e., $\mathbb{P}[C_i \geq X_i] = 1$) has been considered in Example 1. Recall that $\widehat{\Gamma}_n(x) = n^{-1}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$ in this example. The exchangeable bootstrap analog is $\widehat{\Gamma}_n^*(x) = n^{-1}\sum_{i=1}^n W_{i,n}\mathbb{1}(X_i \leq x)$. Assuming f_0 is $\mathfrak{q}$-times differentiable at x for $\mathfrak{q} \geq 1$ with the first $(\mathfrak{q} - 1)$ derivatives vanishing and the $\mathfrak{q}$th derivative positive, Assumptions A and B are easily verified under mild regularity conditions. As a consequence, Theorem 1 implies that the bootstrap-based distributional approximation (3) holds for any $\tilde{M}_{\mathsf{x},n}$ satisfying Assumption C. In particular, $\mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \partial^{\mathfrak{q}} f_0(\mathsf{x})/(\mathfrak{q}+1)!$ in this example, so any pointwise consistent estimator of $\partial^{\mathfrak{q}} f_0(\mathsf{x})$ could be used to estimate $\mathcal{D}_{\mathfrak{q}}(\mathsf{x})$. Alternatively, since Assumption D also holds, $\tilde{M}_{\mathsf{x},n}$ with one of $\tilde{\mathcal{D}}^{\mathtt{MA}}_{\mathfrak{q},n}, \tilde{\mathcal{D}}^{\mathtt{FD}}_{\mathfrak{q},n}, \tilde{\mathcal{D}}^{\mathtt{BR}}_{\mathfrak{q},n}$ can also be used, provided that $\epsilon_n \to 0$ and $n\epsilon_n^{1+2\mathfrak{q}} \to \infty$.

Next, suppose that censoring occurs completely at random; that is, suppose $X_i \perp\!\!\!\perp C_i$. (See Huang and Wellner (1995), and references therein.) In this case, we take $\widehat{\Gamma}_n(x) = 1 - \widehat{S}_n(x)$, where $\widehat{S}_n$ denotes an estimator of the survival function $S_0(x) = \mathbb{P}[X > x]$. For concreteness, let $\widehat{S}_n$ be the Kaplan-Meier estimator. Letting $\widehat{\Gamma}_n^*$ be the natural bootstrap analog of $\widehat{\Gamma}_n$, the conclusions from the previous paragraph remain valid provided that $\mathbb{P}[C_i > c]$ is continuous, $S_0(u_0)\mathbb{P}[C_i > u_0] > 0$, and other regularity conditions hold. Therefore, if $\tilde{M}_{\mathsf{x},n}$ satisfies Assumption C, then the "reshaped" bootstrap estimator $\widetilde{\theta}_n^*(\mathsf{x}) = \partial_-\mathrm{GCM}_{[0,\widehat{u}_n^*]}(\widetilde{\Gamma}_n^*)(\mathsf{x})$ gives a bootstrap-based distributional approximation satifying (9).

Finally, consider the case of censoring at random; that is, suppose $X_i \perp\!\!\!\perp C_i|\mathbf{A}_i$. (See van der

Laan and Robins (2003), Zeng (2004), and references therein.) Now we set

$$\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \widehat{F}_n(x|\mathbf{A}_i) + \widehat{S}_n(x|\mathbf{A}_i)\Big[\frac{\Delta_i \mathbb{1}(\check{X}_i \leq x)}{\widehat{S}_n(\check{X}_i|\mathbf{A}_i)\widehat{G}_n(\check{X}_i|\mathbf{A}_i)} - \int_0^{\min\{\check{X}_i,x\}} \frac{d\widehat{\Lambda}_n(u|\mathbf{A}_i)}{\widehat{S}_n(u|\mathbf{A}_i)\widehat{G}_n(u|\mathbf{A}_i)}\Big],$$

where $\widehat{F}_n(x|\mathbf{A}_i) = 1 - \widehat{S}_n(x|\mathbf{A}_i)$, $\widehat{S}_n(x|\mathbf{A}_i)$ and $\widehat{G}_n(c|\mathbf{A}_i)$ denote preliminary estimates of the conditional survival functions $S_0(x|\mathbf{A}_i) = \mathbb{P}[X_i > x|\mathbf{A}_i]$ and $G_0(c|\mathbf{A}_i) = \mathbb{P}[C_i > c|\mathbf{A}_i]$, respectively, and $\widehat{\Lambda}_n(u|\mathbf{A}_i)$ is the conditional cumulative hazard function that corresponds to $\widehat{S}_n(u|\mathbf{A}_i)$. Letting

$$\widehat{\gamma}_n(x;\mathbf{Z}_i) = \widehat{F}_n(x|\mathbf{A}_i) + \widehat{S}_n(x|\mathbf{A}_i)\Big[\frac{\Delta_i \mathbb{1}(\check{X}_i \leq x)}{\widehat{S}_n(\check{X}_i|\mathbf{A}_i)\widehat{G}_n(\check{X}_i|\mathbf{A}_i)} - \int_0^{\min\{\check{X}_i,x\}} \frac{d\widehat{\Lambda}_n(u|\mathbf{A}_i)}{\widehat{S}_n(u|\mathbf{A}_i)\widehat{G}_n(u|\mathbf{A}_i)}\Big]$$

a bootstrap analog of $\widehat{\Gamma}_n$ is given by $\widehat{\Gamma}^*_n(x) = n^{-1}\sum_{i=1}^n W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i)$, where we employ the original first-step estimates $\widehat{S}_n(x|\mathbf{A}_i)$ and $\widehat{G}_n(c|\mathbf{A}_i)$. This case also fits into the setting of Section 4.2, with the estimator $\widehat{\gamma}_n(x;\mathbf{Z}_i)$ defined above. Under regularity conditions stated in the supplemental appendix and if $\tilde{M}_{\mathsf{x},n}$ is one of the numerical derivative-based estimators discussed in Section 4.1, then the "reshaped" bootstrap estimator $\widetilde{\theta}^*_n(\mathsf{x}) = \partial_-\mathrm{GCM}_{[0,\widehat{u}^*_n]}(\widetilde{\Gamma}^*_n)(\mathsf{x})$ gives a bootstrap-based distributional approximation satifying (9).

5.2 Monotone Regression Estimation

As a second pair of examples, consider the problem of monotone regression estimation, possibly with additional covariate adjustment. We abstract from censoring and assume that $\mathbf{Z}_i = (Y_i, X_i, \mathbf{A}_i)'$, $i = 1, 2, \ldots, n$, is an observed random sample. The parameter of interest is $\theta_0(x) = \mathbb{E}[\mu_0(x|\mathbf{A}_i)]$, where $\mu_0(x|a) = \mathbb{E}[Y_i|X_i = x, \mathbf{A}_i = a]$ and where x is an interior point of I, the support of X. (If there are no covariates $\mathbf{A}$, then $\theta_0(x) = \mu_0(x) = \mathbb{E}[Y_i|X_i = x]$.) It is assumed that θ_0 is non-decreasing on I.

Setting $\Phi_0 = F_0$, the cdf of X, we have $u_0 = 1$ and we therefore set $\widehat{u}_n = \widehat{u}^*_n = 1$. It remains to specify $\widehat{\Gamma}_n, \widehat{\Gamma}^*_n, \widehat{\Phi}_n$ and $\widehat{\Phi}^*_n$.

The classical monotone regression estimator has been considered in Example 2. For that estimator, $\widehat{\Gamma}_n(x) = n^{-1}\sum_{i=1}^n Y_i\mathbb{1}(X_i \leq x)$ and $\widehat{\Phi}_n(x) = n^{-1}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$, bootstrap analogs of which are $\widehat{\Gamma}^*_n(x) = n^{-1}\sum_{i=1}^n W_{i,n}Y_i\mathbb{1}(X_i \leq x)$ and $\widehat{\Phi}^*_n(x) = n^{-1}\sum_{i=1}^n W_{i,n}\mathbb{1}(X_i \leq x)$, respectively. Assuming $f_0(\mathsf{x})$ is positive and that μ_0 is $\mathfrak{q}$-times differentiable at x for $\mathfrak{q} \geq 1$ with the first $(\mathfrak{q}-1)$ derivatives vanishing and the $\mathfrak{q}$th derivative positive, Assumptions A and B are easily verified under mild regularity conditions. As a consequence, Theorem 1 implies that the bootstrap-based distributional approximation (3) holds for any $\tilde{M}_{\mathsf{x},n}$ satisfying Assumption C. In this example, $\mathcal{D}_\mathfrak{q}(\mathsf{x}) = f_0(\mathsf{x})\partial^\mathfrak{q}\mu(\mathsf{x})/(\mathfrak{q}+1)!$, so any pointwise consistent estimators of $f_0(\mathsf{x})$ and $\partial^\mathfrak{q}\mu_0(\mathsf{x})$ could be used to estimate $\mathcal{D}_\mathfrak{q}(\mathsf{x})$. 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. (See

Westling et al. (2020).) We take $\widehat{\Phi}_n(x) = n^{-1}\sum_{i=1}^{n}\mathbb{1}(X_i \leq x)$ and

$$\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i \leq x)\Big[\frac{Y_i - \widehat{\mu}_n(X_i|\mathbf{A}_i)}{\widehat{g}_n(X_i|\mathbf{A}_i)} + \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}_n(X_i|\mathbf{A}_j)\Big],$$

where $\widehat{\mu}_n(x|\mathbf{a})$ and $\widehat{g}_n(x|\mathbf{a})$ are preliminary estimators of $\mu_0(x|\mathbf{a})$ and $g_0(x|\mathbf{a}) = f_0(x|\mathbf{a})/f_0(x)$, respectively, with $f_0(x|\mathbf{a})$ denoting the conditional density of X given $\mathbf{A}$. Bootstrap analogs of $\widehat{\Phi}_n$ and $\widehat{\Gamma}_n$ are given by $\widehat{\Phi}_n^*(x) = \frac{1}{n}\sum_{i=1}^{n}W_{i,n}\mathbb{1}(X_i \leq x)$ and $\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^{n}W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i)$, respectively, where

$$\widehat{\gamma}_n(x;\mathbf{Z}_i) = \mathbb{1}(X_i \leq x)\Big[\frac{Y_i - \widehat{\mu}_n(X_i|\mathbf{A}_i)}{\widehat{g}_n(X_i|\mathbf{A}_i)} + \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}_n(X_i|\mathbf{A}_j)\Big].$$

Under regularity conditions stated in the supplemental appendix and if $\tilde{M}_{\mathsf{x},n}$ is one of the numerical derivative-based estimators discussed in Section 4.1, then the "reshaped" bootstrap estimator $\widetilde{\theta}_n^*(\mathsf{x}) = \partial_-\mathrm{GCM}_{[0,1]}(\widetilde{\Gamma}_n^*)(\mathsf{x})$ gives a bootstrap-based distributional approximation satifying (9).

6 Simulations

We consider the canonical case of monotone density estimation, and employ the simulation setting previously used in Sen et al. (2010). We estimate the monotone Lebesgue density function $\theta_0(\mathsf{x}) = f(\mathsf{x})$ at an interior point $\mathsf{x} = 1$ using a random sample of observations, where three distinct distributions are considered: Model 1 corresponds to $\mathsf{Exponential}(1)$, Model 2 corresponds to $|\mathsf{Normal}(0,1)|$ and Model 3 corresponds to $|\mathcal{T}_3|$. The Monte Carlo experiment employs a sample size $n = 1,000$ with $B = 2,000$ bootstrap replications and $S = 2,000$ simulations, and compares three types of bootstrap-based inference procedures: the standard non-parametric bootstrap, m-out-of-n bootstrap, and our proposed bootstrap-based inference method implemented using two distinct mean function estimators of the form $\tilde{M}^{\mathsf{q}}_{\mathsf{x},n} = \tilde{\mathcal{D}}_{\mathsf{q},n}(\mathsf{x})|v|^2$. More precisely, we consider (i) $\tilde{\mathcal{D}}_{\mathsf{q},n}(\mathsf{x}) = \widehat{\partial^2 f}(\mathsf{x})/2$ an "off-the-shelve" plug-in density estimator (labeled $\tilde{M}^{\mathsf{q},\mathtt{PI}}_{\mathsf{x},n}$), and (ii) $\tilde{\mathcal{D}}_{\mathsf{q},n}(\mathsf{x}) = \tilde{\mathcal{D}}^{\mathtt{FD}}_{\mathsf{q},n}(\mathsf{x})$ (labeled $\tilde{M}^{\mathsf{q},\mathtt{FD}}_{\mathsf{x},n}$).

Table 1 presents the numerical results. We report empirical coverage for nominal 95% confidence intervals and their average interval length. For the case of our proposed procedures, we investigate their performance using both (i) a grid of fixed tuning parameter value (derivative step/bandwidth) and (ii) infeasible and feasible AMSE-optimal choice of tuning parameter. Our proposed bootstrap-based inference method leads to confidence intervals with excellent empirical coverage and average interval length, outperforming both the standard non-parametric bootstrap (which is inconsistent) and the m-out-of-n bootstrap (which is consistent). In particular, in this example, the plug-in method employs an off-the-shelf kernel derivative estimator, which in this case leads to confidence intervals that are very robust (i.e., insensitive) to the choice of bandwidth. Furthermore, when the corresponding feasible off-the-shelf MSE-optimal bandwidth is used, the resulting confidence intervals continue to perform well. Finally, the generic numerical derivative estimator also leads to very good performance of bootstrap-based infeasible and feasible confidence intervals.

A Technical Results

This appendix presents two technical results. First, in Section A.1 we discuss the Generalized Switching Lemma of Westling and Carone (2020, Supplement), demonstrating by example that the lemma is incorrect and presenting a corrected version of it. The corrected lemma includes an additional regularity condition, but also removes other unnecessary conditions. 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. Both lemmas are used in the paper and may be of general interest as well.

A.1 Generalized Switching Lemma

To establish convergence in distribution of generalized Grenander-type estimators, Westling and Carone (2020) relied on the following:

Statement 1 (Lemma 1 in the Supplement to Westling and Carone (2020) restated in our notation)**.** *Let Φ and Γ be functions from a closed interval $I \subseteq \mathbb{R}$ to $\mathbb{R}$, where Φ is non-decreasing and cádlág, Γ and $\Gamma \circ \Phi^{-}$ are lower semi-continuous, and $\{a, b\} \subset \Phi(I) \subset [a, b]$. Let $\theta = \partial_{-}\mathsf{GCM}_{[a,b]}(\Gamma \circ \Phi^{-}) \circ \Phi$. Then, for any $c \in \mathbb{R}$ and $\mathsf{x} \in I$ with $\Phi(\mathsf{x}) \in (a, b)$,*

$$\theta(\mathsf{x}) > c \quad \Longleftrightarrow \quad \sup \operatorname*{argmax}_{v \in I^\star}\{c\Phi(v) - \Gamma(v)\} < \Phi^{-}(\Phi(\mathsf{x})) \tag{A.1}$$

where $I^\star = I \cap \Phi^{-}([a, b]) = \{x \in I : x = \Phi^{-}(u), u \in [a, b]\}$.

As stated, the above lemma is incorrect. There are (at least) two ways in which the conclusion can fail. First, the maximization problem $\max_{v \in I^\star}\{c\Phi(v) - \Gamma(v)\}$ may not have a solution. Furthermore, even when a solution exists, the statement (A.1) can fail. In both cases, the problems can arise when $c < 0$ and are attributable to the fact that $c\Phi$ may not be upper semi-continuous when $c < 0$.

To illustrate what can happen when $c < 0$, let $I = [0, 1]$, $a = 0$, $b = 1$, and define

$$\Phi(v) = \begin{cases} 0.5v & \text{if } 0 \le v < 0.8 \\ v & \text{if } 0.8 \le v \le 1 \end{cases}$$

and

$$\Gamma(v) = \begin{cases} -v & \text{if } 0 \le v < 0.8 \\ 4v - 4 & \text{if } 0.8 \le v \le 1. \end{cases}$$

Then,

$$\Gamma(\Phi^{-}(u)) = \begin{cases} -2u & \text{if } 0 \le u < 0.4 \\ -0.8 & \text{if } 0.4 \le u < 0.8 \\ 4u - 4 & \text{if } 0.8 \le u \le 1 \end{cases}$$

and

$$\theta(x) = \begin{cases} -2 & \text{if } 0 \leq x < 0.8 \\ 0 & \text{if } x = 0.8 \\ 4 & \text{if } 0.8 < x \leq 1. \end{cases}$$

(See also Figure 1.) Note that Φ is non-decreasing and cádlág, Γ and $\Gamma \circ \Phi^{-}$ are continuous, and $\{a, b\} \subset \Phi(I) \subset [a, b]$. As a consequence, the assumptions of Statement 1 are satisfied.

Now consider the case when $c \in (-2, 0)$. Then

$$c\Phi(v) - \Gamma(v) = (1 + c0.5)v\mathbb{1}\{0 \leq v < 0.8\} + (4 + (c-4)v)\mathbb{1}\{0.8 \leq v \leq 1\}$$

and $\operatorname{argmax}_{v\in[0,1]}\{c\Phi(v) - \Gamma(v)\}$ is empty.

Next, consider the case $c = -2$ and $\mathsf{x} = 0.8$. Then

$$c\Phi(v) - \Gamma(v) = 0\mathbb{1}\{0 \leq v < 0.8\} + (4 - 6v)\mathbb{1}\{0.8 \leq v \leq 1\}$$

and $\operatorname{argmax}_{v\in[0,1]}\{c\Phi(v) - \Gamma(v)\} = [0, 0.8)$ so $\sup \operatorname{argmax}_{v\in[0,1]}\{c\Phi(v) - \Gamma(v)\} = 0.8 = \Phi^{-}(\Phi(\mathsf{x}))$, but $\theta(\mathsf{x}) = 0 > c$ and the statement (A.1) is therefore incorrect.

As already mentioned, the failure of the switch relation in the above example is due to $c\Phi - \Gamma$ not necessarily being upper semi-continuous when $c < 0$. The problems highlighted by our example can be avoided by imposing an additional restriction on the function Φ. To be specific, it suffices to assume that the range of Φ is closed. This condition is satisfied in all the applications we consider. On the other hand, it turns out that there is no need to assume lower semi-continuity on the part of Γ. The proof of the following lemma is given in the supplemental appendix.

Lemma A.1 (Generalized Switching Lemma)**.** *Let Φ and Γ be functions from a closed interval $I \subseteq \mathbb{R}$ to $\mathbb{R}$, where Φ is non-decreasing and right continuous, $\Gamma \circ \Phi^{-}$ is lower semi-continuous, $\{a, b\} \subset \Phi(I) \subset [a, b]$, and $\Phi(I)$ is closed. Let $\theta = \partial_{-}\mathsf{GCM}_{[a,b]}(\Gamma \circ \Phi^{-}) \circ \Phi$. Then, for any $c \in \mathbb{R}$ and $\mathsf{x} \in I$ with $\Phi(\mathsf{x}) \in (a, b)$,*

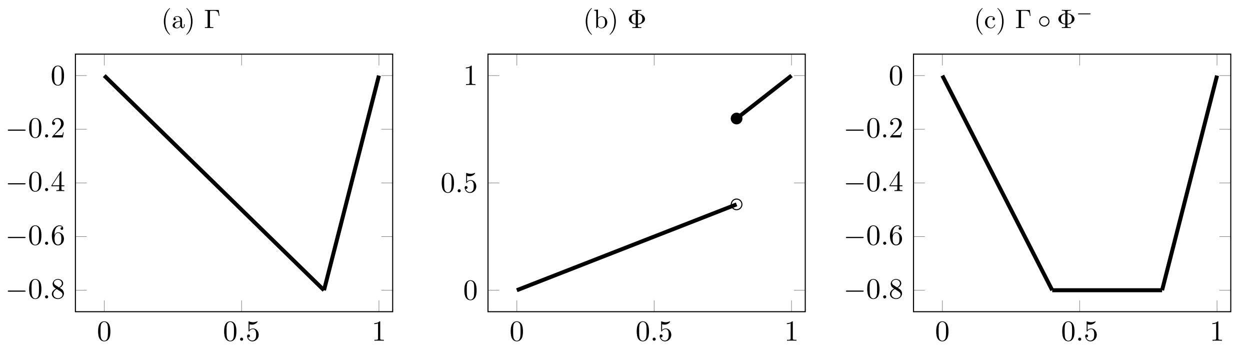
$$\theta(\mathsf{x}) > c \qquad \Longleftrightarrow \qquad \sup \operatorname*{argmax}_{v\in I^{\star}}\{c\Phi(v) - \Gamma(v)\} < \Phi^{-}(\Phi(\mathsf{x}))$$

where $I^{\star} = I \cap \Phi^{-}([a, b]) = \{x \in I : x = \Phi^{-}(u), u \in [a, b]\}$.

A.2 Continuity of argmax

Let $\{\mathbb{G}(v) : v \in \mathbb{R}\}$ be a Gaussian process with mean function μ, covariance kernel $\mathcal{K}$, and continuous sample paths. Under conditions on μ and $\mathcal{K}$ stated below, there exists a unique maximizer of $\mathbb{G}(v)$ over $v \in \mathbb{R}$ with probability one. We complement this known fact with a result showing that the cdf of $\operatorname{argmax}_{v\in\mathbb{R}} \mathbb{G}(v)$ is continuous.

Figure 1: Plots of functions



Assumption A.1. *For every $\tau > 0$ and every $v, v' \in \mathbb{R}$, $\mathcal{K}(v\tau, v'\tau) = \mathcal{K}(v, v')\tau$ and*

$$\mathcal{K}(v + v', v + v') - \mathcal{K}(v + v', v') - \mathcal{K}(v', v + v') + \mathcal{K}(v', v') = \mathcal{K}(v, v).$$

In addition, $\mathcal{K}(1, 1) > 0$, and $\lim_{\delta \downarrow 0} \mathcal{K}(1, \delta)/\sqrt{\delta} = 0$.

Assumption A.2. *For some $c > 1$, $\limsup_{|v| \to \infty} \mu(v)|v|^{-c} = -\infty$.*

Lemma A.2. *Suppose that Assumptions A.1 and A.2 hold. Then $x \mapsto \mathbb{P}[\operatorname{argmax}_{v \in \mathbb{R}}\{\mathbb{G}(v)\} \leq x]$ is continuous.*

Under Assumptions A and B of the main text, $\mathcal{K} \equiv \mathcal{C}_{\mathsf{x}}$ and $\mu \equiv -\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}} + t\mathcal{L}_{\mathsf{x}}$ satisfy Assumptions A.1 and A.2 for any $t \in \mathbb{R}$. As a consequence, it follows from the lemma for any $t \in \mathbb{R}$, the distribution function

$$x \mapsto \mathbb{P}\Big[\operatorname*{argmax}_{v \in \mathbb{R}}\{-\mathcal{G}_{\mathsf{x}}(v) - \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) + t\mathcal{L}_{\mathsf{x}}(v)\} \leq x\Big]$$

is continuous at $x = 0$. We utilize that fact in our proof of (3) and note in passing that most of the existing literature on monotone function estimators seems to implicitly utilize a similar continuity result.

References

Brunk, H. D. (1958), "On the Estimation of Parameters Restricted by Inequalities," *Annals of Mathematical Statistics*, 29, 437–454.

Cattaneo, M. D., Jansson, M., and Nagasawa, K. (2020), "Bootstrap-Based Inference for Cube Root Asymptotics," *Econometrica*, 88, 2203–2219.

Cavaliere, G., and Georgiev, I. (2020), "Inference under Random Limit Bootstrap Measures," *Econometrica*, 88, 2547–2574.

Chernoff, H. (1964), "Estimation of the Mode," *Annals of the Institute of Statistical Mathematics*, 16, 31–41.

Grenander, U. (1956), "On the Theory of Mortality Measurement: Part II," *Scandinavian Actuarial Journal*, 39, 125–153.

Groeneboom, P. (1985), "Estimating a monotone density," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II*, Wadsworth Statist./Probab., pp. 539–555.

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

——— (2018), "Some Developments in the Theory of Shape Constrained Inference," *Statistical Science*, 33, 473–492.

Han, Q., and Kato, K. (2022), "Berry-Esseen Bounds for Chernoff-type Non-standard Asymptotics in Isotonic Regression," *Annals of Applied Probability*, 32, 1459–1498.

Hong, H., and Li, J. (2020), "The Numerical Bootstrap," *Annals of Statistics*, 48, 397–412.

Huang, J., and Wellner, J. A. (1995), "Estimation of a Monotone Density or Monotone hazard under Random Censoring," *Scandinavian Journal of Statistics*, 22, 3–33.

Kim, J., and Pollard, D. (1990), "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

Kosorok, M. R. (2008), "Bootstrapping the Grenander Estimator," in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, Institute of Mathematical Statistics, pp. 282–292.

Lee, S. M. S., and Yang, P. (2020), "Bootstrap Confidence Regions Based on M-Estimators under Nonstandard Conditions," *Annals of Statistics*, 48, 274–299.

Politis, D. N., and Romano, J. P. (1994), "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22, 2031–2050.

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*, Springer Science & Business Media.

Sen, B., Banerjee, M., and Woodroofe, M. (2010), "Inconsistency of Bootstrap: The Grenander Estimator," *Annals of Statistics*, 38, 1953–1977.

Seo, M. H., and Otsu, T. (2018), "Local M-Estimation with Discontinuous Criterion for Dependent and Limited Observations," *Annals of Statistics*, 46, 344–369.

van der Laan, M. J., and Robins, J. M. (2003), *Unified Methods for Censored longitudinal Data and Causality*, Springer Science & Business Media.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.

van der Vaart, A. W., and van der Laan, M. J. (2006), "Estimating a Survival Distribution with Current Status Data and High-Dimensional Covariates," *International Journal of Biostatistics*, 2.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.

Westling, T., and Carone, M. (2020), "A Unified Study of Nonparametric Inference for Monotone Functions," *Annals of Statistics*, 48, 1001–1024.

Westling, T., Gilbert, P., and Carone, M. (2020), "Causal isotonic regression," *Journal of the Royal Statistical Society, Series B*, 82, 719–747.

Zeng, D. (2004), "Estimating Marginal Survival Function by Adjusting for Dependent Censoring using Many Covariates," *Annals of Statistics*, 32, 1533–1555.

Table 1: Simulations, Isotonic Density Estimator, 95% Confidence Intervals.

(a) $n = 1,000$, $S = 2,000$, $B = 2,000$

	Model 1			Model 2			Model 3		
	h, ϵ	Coverage	Length	h, ϵ	Coverage	Length	h, ϵ	Coverage	Length
Standard									
		0.828	0.146		0.808	0.172		0.821	0.155
m-out-of-n									
$m = \lceil n^{1/2} \rceil$		1.000	0.438		0.995	0.495		0.998	0.452
$m = \lceil n^{2/3} \rceil$		0.989	0.314		0.979	0.360		0.989	0.328
$m = \lceil n^{4/5} \rceil$		0.953	0.235		0.937	0.274		0.948	0.248
Plug-in: $\widetilde{Q}_n^{\mathtt{PI}}$									
$0.7 \cdot h_{\mathtt{MSE}}$	0.264	0.955	0.157	0.202	0.947	0.183	0.209	0.957	0.165
$0.8 \cdot h_{\mathtt{MSE}}$	0.302	0.954	0.157	0.231	0.946	0.182	0.239	0.952	0.165
$0.9 \cdot h_{\mathtt{MSE}}$	0.339	0.951	0.156	0.260	0.944	0.181	0.269	0.949	0.164
$1.0 \cdot h_{\mathtt{MSE}}$	0.377	0.949	0.154	0.289	0.941	0.180	0.299	0.948	0.163
$1.1 \cdot h_{\mathtt{MSE}}$	0.415	0.940	0.151	0.318	0.938	0.178	0.329	0.944	0.161
$1.2 \cdot h_{\mathtt{MSE}}$	0.452	0.934	0.147	0.347	0.934	0.176	0.359	0.939	0.158
$1.3 \cdot h_{\mathtt{MSE}}$	0.490	0.922	0.142	0.376	0.928	0.173	0.389	0.935	0.155
$h_{\mathtt{AMSE}}$	0.380	0.949	0.154	0.300	0.940	0.180	0.333	0.943	0.161
$\hat{h}_{\mathtt{AMSE}}$	0.364	0.950	0.155	0.290	0.941	0.180	0.401	0.930	0.154
Num Deriv: $\widetilde{Q}_n^{\mathtt{ND}}$									
$0.7 \cdot \epsilon_{\mathtt{MSE}}$	0.726	0.954	0.158	0.527	0.947	0.183	0.554	0.952	0.165
$0.8 \cdot \epsilon_{\mathtt{MSE}}$	0.830	0.956	0.159	0.602	0.947	0.182	0.633	0.950	0.164
$0.9 \cdot \epsilon_{\mathtt{MSE}}$	0.933	0.956	0.160	0.678	0.944	0.181	0.712	0.949	0.163
$1.0 \cdot \epsilon_{\mathtt{MSE}}$	1.037	0.956	0.159	0.753	0.942	0.180	0.791	0.948	0.162
$1.1 \cdot \epsilon_{\mathtt{MSE}}$	1.141	0.955	0.159	0.828	0.940	0.179	0.870	0.946	0.161
$1.2 \cdot \epsilon_{\mathtt{MSE}}$	1.244	0.956	0.160	0.904	0.936	0.177	0.949	0.943	0.160
$1.3 \cdot \epsilon_{\mathtt{MSE}}$	1.348	0.960	0.163	0.979	0.935	0.176	1.028	0.940	0.159
$\epsilon_{\mathtt{AMSE}}$	0.927	0.956	0.160	0.731	0.943	0.180	0.812	0.948	0.162
$\hat{\epsilon}_{\mathtt{AMSE}}$	0.888	0.956	0.159	0.708	0.943	0.181	0.978	0.942	0.159

Notes:
(i) Panel **Standard** refers to standard non-parametric bootstrap, Panel **m-out-of-n** refers to m-out-of-n non-parametric bootstrap with subsample m, Panel **Plug-in:** $\widetilde{Q}_n^{\mathtt{PI}}$ refers to our proposed bootstrap-based implemented using the example-specific plug-in mean function estimator, and Panel **Num Deriv:** $\widetilde{Q}_n^{\mathtt{ND}}$ refers to our proposed bootstrap-based implemented using the generic numerical derivative mean function estimator.
(ii) Column "h, ϵ" reports tuning parameter value used or average across simulations when estimated, and Columns "Coverage" and "Length" report empirical coverage and average length of bootstrap-based 95% percentile confidence intervals, respectively.
(iii) $h_{\mathtt{MSE}}$ and $\epsilon_{\mathtt{MSE}}$ correspond to the simulation MSE-optimal choices, $h_{\mathtt{AMSE}}$ and $\epsilon_{\mathtt{AMSE}}$ correspond to the AMSE-optimal choices, and $\widehat{h}_{\mathtt{AMSE}}$ and $\widehat{\epsilon}_{\mathtt{AMSE}}$ correspond to the ROT feasible implementation of $\widehat{h}_{\mathtt{AMSE}}$ and $\widehat{\epsilon}_{\mathtt{AMSE}}$ described in the supplemental appendix.