The Annals of Statistics
2011, Vol. 39, No. 3, 1580–1607
DOI: 10.1214/11-AOS874
© Institute of Mathematical Statistics, 2011

CHANGE-POINT IN STOCHASTIC DESIGN REGRESSION AND THE BOOTSTRAP

BY EMILIO SEIJO AND BODHISATTVA SEN[1]

Columbia University

In this paper we study the consistency of different bootstrap procedures for constructing confidence intervals (CIs) for the unique jump discontinuity (change-point) in an otherwise smooth regression function in a stochastic design setting. This problem exhibits nonstandard asymptotics, and we argue that the standard bootstrap procedures in regression fail to provide valid confidence intervals for the change-point. We propose a version of smoothed bootstrap, illustrate its remarkable finite sample performance in our simulation study and prove the consistency of the procedure. The m out of n bootstrap procedure is also considered and shown to be consistent. We also provide sufficient conditions for any bootstrap procedure to be consistent in this scenario.

1. Introduction. Change-point models may arise when a stochastic system is subject to sudden external influences and are encountered in almost every field of science. In the simplest form the model considers a random vector $X = (Y, Z)$ satisfying the following relation:

$$Y = \alpha_0 \mathbf{1}_{Z \le \zeta_0} + \beta_0 \mathbf{1}_{Z > \zeta_0} + \varepsilon, \tag{1}$$

where Z is a continuous random variable, $\alpha_0 \ne \beta_0 \in \mathbb{R}$, $\zeta_0 \in [a, b] \subset \mathbb{R}$ and ε is a continuous random variable, independent of Z with zero expectation and finite variance $\sigma^2 > 0$. The parameter of interest is ζ_0, the change-point.

Despite its simplicity, model (1) captures the inherent "nonstandard" nature of the problem: the least squares estimator of the change-point ζ_0 converges at a rate of n^{-1} to a minimizer of a two-sided, compound Poisson process that depends crucially on the entire error distribution, the marginal density of Z, among other nuisance parameters (see [15], Section 14.5.1, pages 271–277, [17] or [23]). Therefore, it is not practical to use this limiting distribution to build CIs for ζ_0. Bootstrap methods bypass the estimation of nuisance parameters and are generally reliable in $\sqrt{n}$-convergence problems. In this paper we investigate the performance (both theoretically and through simulation) of different bootstrap schemes in building CIs for ζ_0. We hope that the analysis of the bootstrap procedures employed in this

Received September 2010.
[1]Supported by NSF Grant DMS-09-06597.
MSC2010 subject classifications. Primary 62G08, 62G05; secondary 62G20.
Key words and phrases. Argmax continuous mapping theorem, consistency of the bootstrap, m out of n bootstrap, nonstandard asymptotics, semiparametric regression, smoothed bootstrap.

paper will help illustrate the issues that arise when the bootstrap is applied in such nonstandard problems.

The problem of estimating a jump-discontinuity (change-point) in an otherwise smooth curve has been under study for at least the last forty years. More recently, it has been extensively studied in the nonparametric regression and survival analysis literature (see, e.g., [5, 11, 16, 18, 23] and the references therein). Bootstrap techniques have also been applied in many instances in change-point models. Dümbgen [7] proposed asymptotically valid confidence regions for the change-point by inverting bootstrap tests in a one-sample problem. Hůsková and Kirch [13] considered bootstrap CIs for the change-point of the mean in a time series context. A form of parametric bootstrap was used in [16] to estimate the distribution of the estimated change-point in a stochastic design regression model that arises in survival analysis. In a slightly different setting Gijbels, Hall and Kneip [12] suggested a bootstrap procedure for the model (1), but did not give a complete proof of its validity.

Our work goes beyond those cited above as follows: we present strong theoretical and empirical evidence to suggest the *inconsistency* of the two most natural bootstrap procedures in a regression setup—the usual nonparametric bootstrap [i.e., sampling from the empirical cumulative distribution function (ECDF) of (Y, Z), often also called as bootstrapping "pairs"] and the "residual" bootstrap. The bootstrap estimators constructed by these two methods are the smallest maximizers of certain stochastic processes. We show that, conditional on the data, these processes do not have any weak limit in probability. This fact strongly suggests not only inconsistency but also the absence of *any* weak limit for the bootstrap estimators. In the case of the ECDF bootstrap, we also provide an alternative argument for inconsistency via a careful analysis of the *unconditional* behavior of the bootstrap estimator. In addition, we prove that independent sampling from a smooth approximation to the marginal of Z and the centered ECDF of the residuals, and the m out of n bootstrap from the ECDF of (Y, Z) yield asymptotically valid CIs for ζ_0. The finite sample performance of the different bootstrap methods shows the superiority of the proposed smoothed bootstrap procedure. We also develop a series of convergence results which generalize those obtained in [15] to triangular arrays of random vectors and can be used to validate the consistency of *any* bootstrap scheme in this setup. Moreover, in the process of achieving this we develop convergence results for stochastic processes with a three-dimensional parameter which are continuous on the first two arguments and cádlág on the third.

Although we develop our results in the setting of (1), our conclusions have broader implications. They extend immediately to regression functions with parametrically specified models on either side of the change-point (as discussed in Section 7). The *smoothed bootstrap* procedure can also be modified to work in more general nonparametric settings. Gijbels et al. [11], in the second stage of their two-stage procedure to build CI for the change-point in the more general setup of nonparametric regression, localize to a neighborhood of the change-point and reduce the problem to exactly that of (1). Lan et al. [18] consider a two-stage adaptive sampling procedure to estimate the jump discontinuity. The second stage of their method relies on an approximate CI for the change-point, and the bootstrap methods developed in this paper can be immediately used in their context.

The paper is organized in the following manner: in Section 2 we describe the problem in greater detail, introduce the bootstrap schemes and describe the appropriate notion of consistency. In Section 3, we state a series of convergence results that generalize those obtained in [15]. We study the inconsistency of the standard bootstrap methods, including the ECDF and residual bootstraps in Section 4. In Section 5 we prove the consistency of the smoothed and the m out of n bootstrap procedures. We compare the finite sample performance of the different bootstrap methods through a simulation study in Section 6. In Section 7 we discuss the consequences of our analysis to more general change-point regression models. For the sake of brevity we have relegated some technical results to Appendix and to the supplementary paper [25].

2. The problem and the bootstrap schemes. Assume that we are given an i.i.d. sequence of random vectors $\{X_n = (Y_n, Z_n)\}_{n=1}^{\infty}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ having a common distribution $\mathbb{P}$ satisfying (1) for some parameter $\theta_0 := (\alpha_0, \beta_0, \zeta_0) \in \Theta := \mathbb{R}^2 \times [a, b]$. This is a semi-parametric model with an Euclidean parameter θ_0 and two infinite-dimensional parameters—the distributions of Z and ε. We are interested in estimating ζ_0, the change-point. For technical reasons, we will also assume that $\mathbb{P}(|\varepsilon|^3) < \infty$. Here, and in the remainder of the paper, we take the convention that for any probability distribution μ, we will denote the expectation operator by $\mu(\cdot)$. In addition, we suppose that $Z \sim F$ with a uniformly bounded density f on $[a, b]$ such that $\inf_{|z-\zeta_0|\leq\eta} f(z) > \kappa > 0$ for some $\eta > 0$ and that $\mathbb{P}(Z < a) \wedge \mathbb{P}(Z > b) > 0$. For $\theta = (\alpha, \beta, \zeta) \in \Theta$, $x = (y, z) \in \mathbb{R}^2$ write

$$m_\theta(x) := -(y - \alpha\mathbf{1}_{z\leq\zeta} - \beta\mathbf{1}_{z>\zeta})^2, \tag{2}$$

$\mathbb{P}_n$ for the empirical measure defined by $X_1, \ldots, X_n$,

$$M_n(\theta) := \mathbb{P}_n(m_\theta) = -\frac{1}{n}\sum_{i=1}^{n}(Y_i - \alpha\mathbf{1}_{Z_i\leq\zeta} + \beta\mathbf{1}_{Z_i>\zeta})^2$$

and $M(\theta) := \mathbb{P}(m_\theta)$. The function M_n is strictly concave in its first two coordinates but càdlàg (right continuous with left limits) in the third; in fact, piecewise constant with n jumps (w.p. 1). Thus, M_n has unique maximizing values of α and β, but an entire interval of maximizers for ζ. Note that (α, β, ζ) is a maximizer of M_n if $M_n(\alpha, \beta, \zeta) \vee M_n(\alpha, \beta, \zeta^-) = \sup\{M_n(\theta) : \theta \in \Theta\}$. For this reason, we define the *least squares estimator* of θ_0 to be the maximizer of M_n over Θ with the smallest ζ, and denote it by

$$\hat{\theta}_n := (\hat{\alpha}_n, \hat{\beta}_n, \hat{\zeta}_n) = \operatorname*{sargmax}_{\theta\in\Theta}\{M_n(\theta)\},$$

where sargmax stands for the *smallest argmax*. Although our results would have been equally true had we chosen the greatest maximizer (or the mid-point of the interval of maximizers), we use the smallest argmax as most authors use this convention (see, e.g., [15, 18, 23]).

The asymptotic properties of this least squares estimator are well known. It is shown in [15], pages 271–277, that the asymptotic distribution of $n(\hat{\zeta}_n - \zeta_0)$ is that of the smallest argmax of a two-sided compound Poisson process. However, the limiting process depends on the distribution of ε and the value of the density of Z at ζ_0. Thus there is no straightforward way to build CIs for ζ_0 using this limiting distribution. In this connection we investigate the performance of bootstrap procedures for constructing CIs for ζ_0.

2.1. *Bootstrap*. We start with a brief review of the bootstrap. Our approach is similar to those described in page 72 of [27], pages 3–11 of [22] and [26]. Given a sample $\mathbf{W}_n = \{W_1, W_2, \dots, W_n\} \overset{\text{i.i.d.}}{\sim} L$ (unknown), suppose that the distribution function H_n of some random variable $R_n \equiv R_n(\mathbf{W}_n, L)$ is of interest. The bootstrap method can be broken into three simple steps:

(i) Construct an estimator $\hat{L}_n$ of L from $\mathbf{W}_n$.

(ii) Generate $\mathbf{W}_n^* = \{W_1^*, \dots, W_{m_n}^*\} \overset{\text{i.i.d.}}{\sim} \hat{L}_n$ given $\mathbf{W}_n$, where the $(m_n)_{n=1}^{\infty}$ is a sequence of natural numbers set to satisfy suitable regularity conditions.

(iii) Estimate H_n by $\hat{H}_n$, the conditional CDF of $R_n(\mathbf{W}_n^*, \hat{L}_n)$ given $\mathbf{W}_n$.

Let d denote the Prokhorov metric (as defined in (1.1) and (1.2), page 96 in [9]) or any other metric metrizing weak convergence of probability measures (for distributions L defined on $\mathbb{R}$ one could choose, for instance, the Lévy metric as defined in problem 14.5, page 198 of [3]). We say that $\hat{H}_n$ is *weakly consistent* if $d(H_n, \hat{H}_n) \overset{P}{\to} 0$; if H_n has a weak limit H, this is equivalent to $\hat{H}_n$ converging weakly to H in probability. Similarly, $\hat{H}_n$ is *strongly consistent* if $d(H_n, \hat{H}_n) \overset{\text{a.s.}}{\to} 0$.

The choice of $\hat{L}_n$ mostly considered in the literature is the ECDF. Intuitively, an $\hat{L}_n$ that mimics the essential properties (e.g., smoothness) of the underlying distribution L can be expected to perform well. Despite being a good estimator in most situations, the ECDF can fail to capture some properties of L that may be crucial for the problem under consideration. This is especially true for nonstandard problems. In Section 4 we illustrate this phenomenon (the inconsistency of the ECDF bootstrap) when $n(\hat{\zeta}_n - \zeta_0)$ is the random variable of interest.

In our context, a consistent bootstrap procedure must approximate the CDF of $\Delta_n = n(\hat{\zeta}_n - \zeta_0)$ with the conditional CDF of $\Delta_n^* = m_n(\zeta_n^* - \hat{\zeta}_n)$ given the data, where ζ_n^* is the least squares estimator of ζ_0 obtained from the bootstrap sample. In the following we introduce four bootstrap schemes that arise naturally in this problem.

Scheme 1 (*ECDF bootstrap*). Draw a bootstrap sample $(Y_{n,1}^*, Z_{n,1}^*), \ldots, (Y_{n,n}^*, Z_{n,n}^*)$ from the ECDF of $(Y_1, Z_1), \ldots, (Y_n, Z_n)$; probably the most widely used bootstrap scheme.

Scheme 2 (*Bootstrapping residuals*). This is another widely used bootstrap procedure in regression models. We first obtain the residuals

$$\hat{\varepsilon}_{n,j} := Y_j - \hat{\alpha}_n \mathbf{1}_{Z_j \le \hat{\zeta}_n} - \hat{\beta}_n \mathbf{1}_{Z_j > \hat{\zeta}_n} \qquad \text{for } j = 1, \ldots, n$$

from the fitted model. Note that these residuals are not guaranteed to have mean 0, so we work with the centered residuals, $\hat{\varepsilon}_{n,1} - \bar{\varepsilon}_n, \ldots, \hat{\varepsilon}_{n,n} - \bar{\varepsilon}_n$, where $\bar{\varepsilon}_n = \sum_{j=1}^n \hat{\varepsilon}_{n,j}/n$. Letting $\mathbb{P}_n^\varepsilon$ denote the empirical measure of the centered residuals, we obtain the bootstrap sample $(Y_{n,1}^*, Z_1), \ldots, (Y_{n,n}^*, Z_n)$ as:

(1) Sample $\varepsilon_{n,1}^*, \ldots, \varepsilon_{n,n}^*$ independently from $\mathbb{P}_n^\varepsilon$.

(2) Fix the predictors Z_j, $j = 1, \ldots, n$, and define the bootstrapped responses at Z_j as $Y_{n,j}^* = \hat{\alpha}_n \mathbf{1}_{Z_j \le \hat{\zeta}_n} + \hat{\beta}_n \mathbf{1}_{Z_j > \hat{\zeta}_n} + \varepsilon_{n,j}^*$. Compute ζ_n^* from $(Y_{n,1}^*, Z_1), \ldots, (Y_{n,n}^*, Z_n)$.

Scheme 3 (*Smoothed bootstrap*). Notice that in (1), Z is assumed to have a density and it also arises in the limiting distribution of Δ_n. A successful bootstrap scheme must mimic this underlying assumption, and we accomplish this in the following:

(1) Choose an appropriate nonparametric smoothing procedure (e.g., kernel density estimation) to build a distribution $\hat{F}_n$ with a density $\hat{f}_n$ such that $\|\hat{F}_n - F\|_\infty \to 0$ a.s. and $\hat{f}_n \to f$ uniformly on some open interval around ζ_0 w.p. 1, where f is the density of Z.

(2) Get i.i.d. replicates $Z_{n,1}^*, \ldots, Z_{n,n}^*$ from $\hat{F}_n$ and sample, independently, $\varepsilon_{n,1}^*, \ldots, \varepsilon_{n,n}^*$ from $\mathbb{P}_n^\varepsilon$.

(3) Define $Y_{n,j}^* = \hat{\alpha}_n \mathbf{1}_{Z_{n,j}^* \le \hat{\zeta}_n} + \hat{\beta}_n \mathbf{1}_{Z_{n,j}^* > \hat{\zeta}_n} + \varepsilon_{n,j}^*$ for all $j = 1, \ldots, n$.

Scheme 4 (*m out of n bootstrap*). A natural alternative to the usual nonparametric bootstrap (i.e., generating bootstrap samples from the ECDF) considered widely in nonregular problems is to use the m out of n bootstrap. We choose a nondecreasing sequence of natural numbers $\{m_n\}_{n=1}^\infty$ such that $m_n = o(n)$ and $m_n \to \infty$ and generate the bootstrap sample $(Y_{n,1}^*, Z_{n,1}^*), \ldots, (Y_{n,m_n}^*, Z_{n,m_n}^*)$ from the ECDF of $(Y_1, Z_1), \ldots, (Y_n, Z_n)$.

We will use the framework established by our convergence theorems in Section 3 to prove that schemes 3 and 4 above yield *consistent* bootstrap procedures for building CIs for ζ_0. We will also give strong empirical and theoretical evidence for the *inconsistency* of schemes 1 and 2. Note that schemes 1 and 2 are the two most widely used resampling techniques in regression models (see pages 35 and 36 of [8]; also see [10] and [29]). Thus in this change-point scenario, a typical nonstandard problem, we see that the two standard bootstrap approaches fail. The failure of the usual bootstrap methods in nonstandard situations is not new and has been investigated in the context of M-estimation problems in [4] and in situations giving rise to $n^{1/3}$ asymptotics in [1] and [26]. But the change-point problem addressed in this paper is indeed quite different from the nonstandard problems considered by the above authors—one key distinction being that compound Poisson processes, as opposed to Gaussian processes, form the backbone of the asymptotic distributions of the estimators and thus demand an independent investigation.

We will also see later that the performance of scheme 3 clearly dominates that of the m out of n bootstrap procedure (scheme 4), the general recipe proposed in situations where the usual bootstrap does not work (see [19] for applications of the m out of n bootstrap procedure in some nonstandard problems). Also note that the performance of the m out of n bootstrap scheme crucially depends on m (see, e.g., [2]) and the choice of this tuning parameter is tricky in applications.

3. A uniform convergence result. In this section we generalize the results obtained in [15], pages 271–277, to a triangular array of random variables. This generalization will help us analyze the asymptotic properties of the bootstrap estimators (to be introduced in Section 4). Conditioned on the data, bootstrap samples can be embedded in a triangular array of random variables, with the nth row being generated from a distribution (built from the first n data points) that approximates the data-generating mechanism. With this in mind, we derive asymptotic results for a general triangular array whose row-distributions satisfy certain regularity conditions. Due to space constraints, we only state the main results; complete proofs can be found in Seijo and Sen [25], the longer version of this paper.

Consider the triangular array $\{X_{n,k} = (Y_{n,k}, Z_{n,k})\}_{1\le k\le m_n}^{n\in\mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, where $(m_n)_{n=1}^{\infty}$ is a nondecreasing sequence of natural numbers such that $m_n \to \infty$. Throughout the paper we denote by $\mathbf{E}$ the expectation operator with respect to $\mathbf{P}$. Furthermore, assume that for each $n \in \mathbb{N}$, $(X_{n,1}, \dots, X_{n,m_n})$ constitutes a random sample from an arbitrary bivariate distribution $\mathbb{Q}_n$ with $\mathbb{Q}_n(Y_{n,1}^2) < \infty$ and let $M_n(\theta) := \mathbb{Q}_n(m_\theta)$ for all $\theta \in \Theta$, where m_θ is defined in (2). Let $\mathbb{P}$ be a bivariate distribution satisfying (1). Recall that $M(\theta) := \mathbb{P}(m_\theta)$ and $\theta_0 := \operatorname{sargmax} M(\theta)$. Let $\theta_n = (\alpha_n, \beta_n, \zeta_n)$ be given by $\theta_n := \operatorname{sargmax}_{\theta\in\Theta}\{\mathbb{Q}_n(m_\theta)\}$. Note that $\mathbb{Q}_n$ need not satisfy model (1) with $(\alpha_n, \beta_n, \zeta_n)$. The existence of θ_n is guaranteed as $\mathbb{Q}_n(m_\theta)$ is a quadratic function in α and β (for a fixed ζ) and bounded and cádlág as a function in ζ. For each n, let $\mathbb{P}_n^*$ be the empirical measure generated by the random sample $(X_{n,1}, \dots, X_{n,m_n})$, and define the least squares estimator $\theta_n^* = (\alpha_n^*, \beta_n^*, \zeta_n^*) \in \Theta$ to be the smallest argmax of $M_n^*(\theta) := \mathbb{P}_n^*(m_\theta)$. If Q is a signed Borel measure on $\mathbb{R}^2$ and $\mathscr{F}$ is a class of complex-valued functions defined on $\mathbb{R}^2$, write $\|Q\|_{\mathscr{F}} := \sup\{|Q(f)| : f \in \mathscr{F}\}$.

If $g : K \subset \mathbb{R}^3 \to \mathbb{R}$ is a bounded function, write $\|g\|_K := \sup_{x \in K} |g(x)|$. Also, for $(z, y) \in \mathbb{R}^2$ and $n \in \mathbb{N}$ we write

$$\tilde{\varepsilon}_n := \tilde{\varepsilon}_n(z, y) = y - \alpha_n \mathbf{1}_{z \le \zeta_n} - \beta_n \mathbf{1}_{z > \zeta_n}.$$

Let $M > 0$ be such that $|\alpha_n| \le M$ for all n. We define the following three classes of functions from $\mathbb{R}^2$ into $\mathbb{R}$:

$$\mathcal{F} := \{\mathbf{1}_I(z) : I \subset \mathbb{R} \text{ is an interval}\};$$
$$\mathcal{G} := \{y f(z) : f \in \mathcal{F}\} \cup \{|y + \alpha| f(z) : f \in \mathcal{F}, |\alpha| \le M\};$$
$$\mathcal{H} := \{y^2 f(z) : f \in \mathcal{F}\}.$$

In what follows, we will derive conditions on the distributions $\mathbb{Q}_n$ that will guarantee consistency and weak convergence of θ_n^*.

3.1. *Consistency and the rate of convergence.* We provide first a consistency result for the least squares estimator, whose proof can be found in Section A.2.1 of [25]. To this end, we consider the following set of assumptions:

(I) $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \to 0$;
(II) $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{G}} \to 0$;
(III) $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{H}} \to 0$;
(IV) $\theta_n \to \theta_0$.

PROPOSITION 3.1. *Assume that* (I)–(IV) *hold. Then*, $\theta_n^* \xrightarrow{\mathbf{P}} \theta_0$.

To guarantee the right rate of convergence, we need to assume stronger regularity conditions. In addition to (I)–(IV), we require the following:

(V) There are $\eta, \rho, L > 0$ with the property that for any $\delta \in (0, \eta)$, there is $N > 0$ such that the following inequalities hold for any $n \ge N$:

$$\inf_{1/\sqrt{m_n} \le |\zeta - \zeta_n| < \delta^2} \left\{ \frac{1}{|\zeta - \zeta_n|} \mathbb{Q}_n(\mathbf{1}_{\zeta \wedge \zeta_n < Z \le \zeta \vee \zeta_n}) \right\} > \rho, \tag{3}$$

$$\sup_{|\zeta - \zeta_n| < \delta^2} \{|\mathbb{Q}_n(\tilde{\varepsilon}_n \mathbf{1}_{\zeta \wedge \zeta_n < Z \le \zeta \vee \zeta_n})|\} \le \frac{L\delta}{\sqrt{m_n}}, \tag{4}$$

$$\sup_{|\zeta - \zeta_n| < \delta^2} \{|\mathbb{Q}_n(\tilde{\varepsilon}_n \mathbf{1}_{Z \le \zeta \wedge \zeta_n})| + |\mathbb{Q}_n(\tilde{\varepsilon}_n \mathbf{1}_{Z > \zeta \vee \zeta_n})|\} \le \frac{L}{\sqrt{m_n}}. \tag{5}$$

We would like to point out some facts about (V). It must be noted that (4) and (5) automatically hold in the case where Z and $\tilde{\varepsilon}_n$ are independent under $\mathbb{Q}_n$ with $\mathbb{Q}_n(\tilde{\varepsilon}_n) = 0$. Also, (3) is easily seen to hold when the Z's, under $\mathbb{Q}_n$, have densities f_n converging uniformly to f in some neighborhood of ζ_0, where f is the density of Z under $\mathbb{P}$, a consequence of the classical mean value theorem of calculus.

PROPOSITION 3.2. *Assume that* (I)–(V) *hold. Then* $\sqrt{m_n}(\alpha_n^* - \alpha_n) = O_{\mathbf{P}}(1)$, $\sqrt{m_n}(\beta_n^* - \beta_n) = O_{\mathbf{P}}(1)$ *and* $m_n(\zeta_n^* - \zeta_n) = O_{\mathbf{P}}(1)$.

The proof of the above proposition can be found in Section A.2.2 of [25].

3.2. *Weak convergence and asymptotic distribution.* We start with some additional sets of assumptions:

(VI) For any function $\psi : \mathbb{R} \to \mathbb{C}$ which is either of the form $\psi(x) = e^{i\xi x}$ for some $\xi \in \mathbb{R}$ or defined by $\psi(x) = |x|^p$ for $p = 1, 2$, we have

$$m_n \mathbb{Q}_n(\psi(\tilde{\varepsilon}_n)\mathbf{1}_{\zeta_n - \delta/m_n < Z \le \zeta_n + \eta/m_n}) \to f(\zeta_0)(\delta + \eta)\mathbb{P}(\psi(\varepsilon)) \qquad \text{for all } \eta, \delta > 0.$$

(VII) $\sqrt{m_n}\mathbb{Q}_n(\tilde{\varepsilon}_n \mathbf{1}_{Z \le \zeta_n}) \to 0$ and $\sqrt{m_n}\mathbb{Q}_n(\tilde{\varepsilon}_n \mathbf{1}_{Z > \zeta_n}) \to 0$.

(VIII) $\overline{\lim}_{n\to\infty} \mathbb{Q}_n(|\tilde{\varepsilon}_n|^3) < \infty$.

For $h = (h_1, h_2, h_3) \in \mathbb{R}^3$, let $\vartheta_{n,h} := \theta_n + (\frac{h_1}{\sqrt{m_n}}, \frac{h_2}{\sqrt{m_n}}, \frac{h_3}{m_n})$ and

$$\hat{E}_n(h) := m_n \mathbb{P}_n^*[m_{\vartheta_{n,h}} - m_{\theta_n}].$$

We derive the asymptotic distribution of the process $\hat{E}_n$ and then apply continuous mapping techniques to obtain the limiting distribution of

$$h_n^* := \operatorname*{sargmin}_{h \in \mathbb{R}^3} \hat{E}_n(h) = \bigl(\sqrt{m_n}(\alpha_n^* - \alpha_n), \sqrt{m_n}(\beta_n^* - \beta_n), m_n(\zeta_n^* - \zeta_n)\bigr).$$

For any given compact rectangle $K \subset \mathbb{R}^3$ we will consider these stochastic processes as random elements in the space $\mathcal{D}_K$ of all functions $W : K \to \mathbb{R}$ having "quadrant limits" (as defined in [21]), being continuous from above (again, in the terminology of [21]) and such that $W(\cdot, \cdot, \zeta)$ is continuous for all ζ and $W(\alpha, \beta, \cdot)$ is càdlàg (right continuous having left limits) for all (α, β). For any compact interval $I \subset \mathbb{R}$ let $\Lambda_I = \{\lambda : I \to I | \lambda \text{ is strictly increasing, surjective and continuous}\}$ and write $\|\lambda\| := \sup_{s \ne t \in I} |\log \frac{\lambda(s) - \lambda(t)}{s - t}|$. Then, for any set of the form $K = A \times I$ with $A \subset \mathbb{R}^2$ define the Skorohod topology as the topology given by the metric

$$d_K(\Psi, \Gamma) := \inf_{\lambda \in \Lambda_I} \Bigl\{ \sup_{(\alpha,\beta,\zeta) \in K} \{|\Psi(\alpha, \beta, \zeta) - \Gamma(\alpha, \beta, \lambda(\zeta))|\} + \|\lambda\| \Bigr\}$$

for $\Gamma, \Psi \in \mathcal{D}_K$. Endowed with this metric, $\mathcal{D}_K$ becomes a Polish space (it is a closed subspace of the Polish spaces D_k defined in [21]) and thus the existence of conditional probability distributions for its random elements is ensured (see Theorem 10.2.2 in page 345 of [6]).

Before stating the convergence result, we need to make the following definitions: let $\mathbf{Z}_1 \sim \mathbf{N}(0, \sigma^2 \mathbb{P}(Z \le \zeta_0))$ and $\mathbf{Z}_2 \sim \mathbf{N}(0, \sigma^2 \mathbb{P}(Z > \zeta_0))$ be two independent normal random variables; ν_1 and ν_2 be, respectively, left-continuous and right-continuous, homogeneous Poisson processes with rate $f(\zeta_0) > 0$; $\mathbf{u} = (u_n)_{n=1}^\infty$ and $\mathbf{v} = (v_n)_{n=1}^\infty$ two sequences of i.i.d. random variables having the same distribution as ε under $\mathbb{P}$. Assume, in addition, that $\mathbf{Z}_1$, $\mathbf{Z}_2$, ν_1, ν_2, $\mathbf{v}$ and $\mathbf{u}$ are all mutually independent. Then, define the process $\Xi = (\Xi^{(1)}, \dots, \Xi^{(6)})'$ as

$$\Xi(t) := \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \nu_1(-t)\mathbf{1}_{t<0} \\ \sum_{0<j\le\nu_1(-t)} v_j \mathbf{1}_{t<0} \\ \nu_2(t)\mathbf{1}_{t\ge 0} \\ \sum_{0<j\le\nu_2(t)} u_j \mathbf{1}_{t\ge 0} \end{pmatrix}, \tag{6}$$

and let E^* be given by

$$\begin{aligned} E^*(h) :=\ & 2h_1\Xi^{(1)}(h_3) - h_1^2\mathbb{P}(Z \le \zeta_0) + 2h_2\Xi^{(2)}(h_3) - h_2^2\mathbb{P}(Z > \zeta_0) \\ & + 2(\beta_0 - \alpha_0)\Xi^{(4)}(h_3) - (\alpha_0 - \beta_0)^2\Xi^{(3)}(h_3) \\ & + 2(\alpha_0 - \beta_0)\Xi^{(6)}(h_3) - (\alpha_0 - \beta_0)^2\Xi^{(5)}(h_3) \end{aligned} \tag{7}$$

for $h = (h_1, h_2, h_3) \in \mathbb{R}^3$.

The next lemma describes the distribution of the smallest argmax of E^*. A proof can be found in Section A.2.6 of [25].

LEMMA 3.1. *Consider the process E^* defined in* (7). *Then, for almost every sample path of E^*, $\phi^* = (\phi_1^*, \phi_2^*, \phi_3^*)' := \operatorname{sargmax}_{h\in\mathbb{R}^3}\{E^*(h)\}$ is well defined. Moreover, ϕ_1^*, ϕ_2^* and ϕ_3^* are independent, and ϕ_1^* and ϕ_2^* are distributed as normal random variables with mean* 0 *and variances $\sigma^2/\mathbb{P}(Z \le \zeta_0)$ and $\sigma^2/\mathbb{P}(Z > \zeta_0)$, respectively.*

We are now in a position to state the main convergence result of this section. Its proof is slightly long and involved, so due to space constraints we only give a sketch of the main argument in Section A.1. A complete proof can be found in Section 3.2 of [25].

PROPOSITION 3.3. *If conditions* (I)–(VIII) *hold, then*:

(i) *For any compact rectangle $K \subset \mathbb{R}^3$, $\hat{E}_n \rightsquigarrow E^*$ on $\mathcal{D}_K$, where $\rightsquigarrow$ denotes weak convergence.*

(ii)

$$\begin{aligned} h_n^* &= \bigl(\sqrt{m_n}(\alpha_n^* - \alpha_n), \sqrt{m_n}(\beta_n^* - \beta_n), m_n(\zeta_n^* - \zeta_n)\bigr)' \\ &\rightsquigarrow \operatorname*{sargmax}_{h\in\mathbb{R}^3}\{E^*(h)\}. \end{aligned}$$

If we take $\mathbb{Q}_n = \mathbb{P}$ and $m_n = n$ for all $n \in \mathbb{N}$, it is easily seen that $\theta_n = \theta_0$, and conditions (I)–(VIII) hold. Hence, we immediately get the following corollary.

COROLLARY 3.1. *For the least squares estimators* $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\zeta}_n)$ *based on an i.i.d. sequence* $(X_n)_{n=1}^{\infty}$ *satisfying* (1) *we have*

$$\big(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), n(\hat{\zeta}_n - \zeta_0)\big)' \rightsquigarrow \operatorname*{sargmax}_{h \in \mathbb{R}^3}\{E^*(h)\}.$$

4. Inconsistency of the bootstrap. In this section we argue the inconsistency of the two most common bootstrap procedures in regression: the ECDF bootstrap (scheme 1) and the residual bootstrap (scheme 2). In fact, we provide two arguments to suggest that the standard ECDF bootstrap is inconsistent, and one argument to indicate that is the same for the standard residual based bootstrap. We will show that the bootstrap estimators from both of these schemes are the smallest maximizers of stochastic processes that, conditional on the data, have *no weak limit in probability*. This suggests not only that the schemes produce inconsistent inference, but also that the estimators have no weak limit in probability. In the case of the ECDF bootstrap, we will also support our claim of inconsistency with a careful analysis of the *unconditional behavior* of $\tilde{\Delta}_n^* := (\sqrt{n}(\alpha_n^* - \alpha_0), \sqrt{n}(\beta_n^* - \beta_0), n(\zeta_n^* - \zeta_0))$. Based on the approach used in [14], if consistency holds for the ECDF bootstrap, then the unconditional asymptotic variance of $n(\zeta_n^* - \zeta_0)$ must be twice that of $n(\hat{\zeta}_n - \zeta_0)$. We derive the asymptotic unconditional distributions of $n(\zeta_n^* - \zeta_0)$ and $n(\hat{\zeta}_n - \zeta_0)$ and compute their variances via simulation, as analytic expressions are not available, to show that the former is not twice the latter.

Recall the notation and definitions in the beginning of Section 2. In particular, note that we have i.i.d. random vectors $\{X_n = (Y_n, Z_n)\}_{n=1}^{\infty}$ from (1) with parameter θ_0 defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, and let $\mathbb{P}_n$ be the empirical distribution of the first n data points. We denote by $\mathfrak{X} = \sigma((X_n)_{n=1}^{\infty})$ the σ-algebra generated by the sequence $(X_n)_{n=1}^{\infty}$ and write $\mathbf{P}_{\mathfrak{X}}(\cdot) = \mathbf{P}(\cdot|\mathfrak{X})$ and $\mathbf{E}_{\mathfrak{X}}(\cdot) = \mathbf{E}(\cdot|\mathfrak{X})$. Let $(\mathbb{X}, d)$ be a metric space, and consider the $\mathbb{X}$-valued random elements V and $(V_n)_{n=1}^{\infty}$ defined on $(\Omega, \mathcal{A}, \mathbf{P})$. We say that V_n converges conditionally in probability to V, almost surely, and write $V_n \xrightarrow[\text{a.s.}]{\mathbf{P}_{\mathfrak{X}}} V$, if

$$\mathbf{P}_{\mathfrak{X}}\big(d(V_n, V) > \varepsilon\big) \xrightarrow{\text{a.s.}} 0 \qquad \forall \varepsilon > 0. \tag{8}$$

Similarly, we write $V_n \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathfrak{X}}} V$ and say that V_n converges conditionally in probability to V, in probability, if the left-hand side of (8) converges in probability to 0.

4.1. *Scheme* 1 (*bootstrapping from the ECDF*). Consider the notation and definitions of Section 2.1. To translate this scheme into the framework of Propositions 3.1, 3.2 and 3.3, we set $m_n = n$, $\mathbb{Q}_n = \mathbb{P}_n$ and consider the triangular array $\{X_{n,k}^* = (Y_{n,k}^*, Z_{n,k}^*)\}_{1 \le k \le n}^{n \in \mathbb{N}}$. Moreover, from Lemma 4.1 in [25] we know that $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$, so we can also take $\theta_n = \hat{\theta}_n$. We first prove that the bootstrapped estimators converge conditionally in probability to the true value of the parameters, almost surely.

PROPOSITION 4.1. *For the ECDF bootstrap, we have* $\theta_n^* \xrightarrow[a.s.]{\mathbf{P}_{\mathfrak{X}}} \theta_0$.

PROOF. Since Y has a second moment under $\mathbb{P}$, it is straightforward to see that $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ are VC-subgraph classes with integrable envelopes 1, $|Y| + M$ and Y^2, respectively. It follows that all these classes are Glivenko–Cantelli, and therefore conditions (I)–(III) hold w.p. 1. Also, note that $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$ implies that condition (IV) holds a.s. The result then follows from Proposition 3.1. □

It is evident that condition (VI) does not hold in this situation as we know that

$$n\mathbb{P}_n\left(\zeta_0 - \frac{\eta}{n} < Z \le \zeta_0 + \frac{\delta}{n}\right) \rightsquigarrow \text{Poisson}\big(f(\zeta_0)(\delta + \eta)\big). \tag{9}$$

Hence, we cannot use Proposition 3.3 to derive the limit behavior of h_n^*.

We will now argue that, conditional on the data, $\hat{E}_n$ *does not have any weak limit* in probability. This statement should be thought in terms of the Prokhorov metric (or any other metric metrizing weak convergence on $\mathcal{D}_K$). If we denote by ρ_K the Prokhorov metric on the space of probability measures on $\mathcal{D}_K$ and by μ_n the conditional distribution of $\hat{E}_n$ given $\mathfrak{X}$, to say that $(\hat{E}_n)_{n=1}^\infty$ has no weak limit in probability means that there is no probability measure μ defined on $\mathcal{D}_K$ such that $\rho_K(\mu_n, \mu) \xrightarrow{\mathbf{P}} 0$. There is also the apparent possibility that μ_n could converge to a random limit, but as the ECDF is invariant under permutations, an application of the Hewitt–Savage zero–one law (see page 496 of [3]) would rule out this possibility (see page 1961 of [26]).

The following lemma (proved in Section A.2) will help us show that the (conditional) characteristic functions corresponding to the finite-dimensional distributions of $\hat{E}_n$ fail to have a limit in probability, which would, in particular, imply that $\hat{E}_n$ does not have a weak limit in probability.

LEMMA 4.1. *The following statements hold*:

(i) *For any two real numbers $s < t$, $\{n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \le \zeta_0 + \frac{t}{n})\}_{n=1}^\infty$ does not converge in probability.*

(ii) *There is $h_* > 0$ such that for any $h \ge h_*$, the sequences $\{n\mathbb{P}_n(\hat{\zeta}_n < Z \le \hat{\zeta}_n + \frac{h}{n})\}_{n=1}^\infty$ and $\{n\mathbb{P}_n(\hat{\zeta}_n - \frac{h}{n} < Z \le \hat{\zeta}_n)\}_{n=1}^\infty$ do not converge in probability.*

(iii) *For any two real numbers $s < t$ and any measurable function $\phi : \mathbb{R} \to \mathbb{R}$, $\{n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\zeta_0+s/n<Z\le\zeta_0+t/n})\}_{n=1}^\infty$ does not converge in probability.*

(iv) *Let ϕ be a measurable function which is either nonnegative or nonpositive and such that $\phi(\varepsilon + \alpha_0)$ and $\phi(\varepsilon + \beta_0)$ are nonconstant random variables with finite second moment. Then, there is $h_* > 0$ such that for any $h \ge h_*$, $\{n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\hat{\zeta}_n<Z\le\hat{\zeta}_n+h/n})\}_{n=1}^\infty$ and $\{n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\hat{\zeta}_n-h/n<Z\le\hat{\zeta}_n})\}_{n=1}^\infty$ do not converge in probability.*

With the aid of Lemma 4.1 we are now able to state our main result.

LEMMA 4.2. *There is a compact rectangle $K \subset \mathbb{R}^3$ such that the conditional distribution of $\hat{E}_n$ given $\mathfrak{X}$ does not have a weak limit in probability in $\mathcal{D}_K$.*

PROOF. It is enough to show that there is some h_3 such that $\hat{E}_n(0,0,h_3)$ does not converge in distribution, conditional on the data, in probability. For $h_3 > 0$, a simplification [using (15)] yields

$$\hat{E}_n(0,0,h_3) = (\hat{\alpha}_n - \hat{\beta}_n)\big(n\mathbb{P}_n^*[(2\tilde{\varepsilon}_n - \hat{\alpha}_n + \hat{\beta}_n)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}]\big).$$

Since $\hat{\alpha}_n - \hat{\beta}_n \xrightarrow{\text{a.s.}} \alpha_0 - \beta_0 \neq 0$ we see that $\hat{E}_n(0,0,h_3)$ will converge weakly in probability iff $\Lambda_n := n\mathbb{P}_n^*[(2\tilde{\varepsilon}_n - \hat{\alpha}_n + \hat{\beta}_n)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}]$ converges weakly in probability.

The conditional characteristic function of Λ_n is given by

$$\mathbf{E}_{\mathfrak{X}}(e^{i\xi\Lambda_n}) = \left(1 + \frac{1}{n} n\mathbb{P}_n\big((e^{i\xi(2\tilde{\varepsilon}_n + \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}\big)\right)^n, \tag{10}$$

which converges in probability iff so does $n\mathbb{P}_n((e^{i\xi(2\tilde{\varepsilon}_n + \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n})$. But note that

$$\begin{aligned} &n\mathbb{P}_n\big((e^{i\xi(2\tilde{\varepsilon}_n + \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}\big) \\ &\qquad = n\mathbb{P}_n\big((e^{i\xi(2Y - \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}\big). \end{aligned}$$

It is easily seen that (9) and the fact that $n(\hat{\zeta}_n - \zeta_0) = O_{\mathbf{P}}(1)$ imply that $n\mathbb{P}_n(\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}) = O_{\mathbf{P}}(1)$. Notice that

$$\begin{aligned} &\big|n\mathbb{P}_n\big((e^{i\xi(2Y - \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}\big) \\ &\quad - n\mathbb{P}_n\big((e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n}\big)\big| \\ &\qquad \le n\mathbb{P}_n(\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n})(|\hat{\alpha}_n - \alpha_0| + |\hat{\beta}_n - \beta_0|)|\xi| \xrightarrow{\mathbf{P}} 0. \end{aligned}$$

Thus (10) has a limit in probability iff $n\mathbb{P}_n((e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n})$ has a limit in probability. But a necessary condition for the latter to happen is that its real part, $n\mathbb{P}_n(\mathrm{Re}(e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n})$, converges in probability. Since $\mathrm{Re}(e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1) \le 0$ we can conclude from (iv) of Lemma 4.1 that $n\mathbb{P}_n(\mathrm{Re}(e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n})$ does not converge in probability for all $h_3 \ge h_*$ for some $h_* > 0$ large enough. This in turn implies that, for all $h_3 \ge h_*$, the conditional characteristic function in (10) does not converge in probability, and hence $\hat{E}_n(0,0,h_3)$ has no weak limit in probability. □

Note that $(\sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n), n(\zeta_n^* - \hat{\zeta}_n)) = \operatorname{sargmax}_{h\in\mathbb{R}^3}\{\hat{E}_n(h)\}$. Thus, the fact that the sequence $(\hat{E}_n)_{n=1}^\infty$ does not have a weak limit in probability makes the existence of a weak limit in probability for $n(\zeta_n^* - \hat{\zeta}_n)$ very unlikely. A complete proof this statement may be complicated because the smallest argmax functional is nonlinear, and $\hat{E}_n$ depends on h_3 through indicator functions that do not converge in the limit (see Lemma 4.1). Due to these difficulties we take an alternative approach to argue the inconsistency of the ECDF bootstrap.

REMARK. It must be noted in this connection that the bootstrap scheme estimates the distribution of $(\sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n))$ correctly, and in fact, valid bootstrap based inference can be conducted to obtain CIs for α_0 and β_0. This follows from the fact that, asymptotically, the maximizers of $\hat{E}_n(\cdot,\cdot,h_3)$ do not depend on h_3.

Our next approach to arguing the inconsistency is similar to that of [14] and relies on the asymptotic *unconditional* behavior of

$$\tilde{\Delta}_n^* := \big(\sqrt{n}(\alpha_n^* - \alpha_0), \sqrt{n}(\beta_n^* - \beta_0), n(\zeta_n^* - \zeta_0)\big).$$

For $h \in \mathbb{R}^3$, we write $\tilde{\vartheta}_{n,h} := \theta_0 + (\frac{h_1}{\sqrt{n}}, \frac{h_2}{\sqrt{n}}, \frac{h_3}{n})$ and $\tilde{E}_n(h) := n\mathbb{P}_n^*[m_{\tilde{\vartheta}_{n,h}} - m_{\theta_0}]$. This corresponds to centering the objective function around θ_0.

In what follows we will describe the limiting distribution of $\tilde{E}_n$. We start by introducing some notation. Recall the definitions of the random elements $\mathbf{Z}_1$, $\mathbf{Z}_2$, ν_1, ν_2, $\mathbf{u}$ and $\mathbf{v}$ as in the discussion preceding (6). Also, let $\tau = (\tau_n)_{n=1}^\infty$ and $\kappa = (\kappa_n)_{n=1}^\infty$ be two sequences of i.i.d. Poisson(1) random variables. Assume, in addition, that $\mathbf{Z}_1$, $\mathbf{Z}_2$, ν_1, ν_2, $\mathbf{v}$, $\mathbf{u}$, τ and κ are all mutually independent. Then define the process $\tilde{\Xi} = (\tilde{\Xi}^{(1)}, \dots, \tilde{\Xi}^{(6)})'$ as

$$\begin{aligned}\tilde{\Xi}(t) := \Bigg(\mathbf{Z}_1, \mathbf{Z}_2, &\sum_{0<j\le\nu_1(-t)} \kappa_j \mathbf{1}_{t<0}, \sum_{0<j\le\nu_1(-t)} v_j\kappa_j\mathbf{1}_{t<0},\\ &\sum_{0<j\le\nu_2(t)} \tau_j\mathbf{1}_{t\ge0}, \sum_{0<j\le\nu_2(t)} u_j\tau_j\mathbf{1}_{t\ge0}\Bigg)'\end{aligned}$$

for $t \in \mathbb{R}$, and let $\tilde{E}^*$ be given by

$$\begin{aligned}\tilde{E}^*(h) = {}& 2h_1\tilde{\Xi}^{(1)}(h_3) - h_1^2\mathbb{P}(Z \le \zeta_0) + 2h_2\tilde{\Xi}^{(2)}(h_3) - h_2^2\mathbb{P}(Z > \zeta_0)\\ &+ 2(\beta_0 - \alpha_0)\tilde{\Xi}^{(4)}(h_3) - (\alpha_0 - \beta_0)^2\tilde{\Xi}^{(3)}(h_3)\\ &+ 2(\alpha_0 - \beta_0)\tilde{\Xi}^{(6)}(h_3) - (\alpha_0 - \beta_0)^2\tilde{\Xi}^{(5)}(h_3)\end{aligned}$$

for $h = (h_1, h_2, h_3) \in \mathbb{R}^3$.

Lemma 4.3 (proved in Section A.2.11 of [25]) now states the asymptotic distribution of $\tilde{E}_n$ and that of $n(\zeta_n^* - \zeta_0)$.

TABLE 1
The (unconditional) asymptotic variance of $n(\zeta_n^* - \zeta_0)$ is not twice that of $n(\hat{\zeta}_n - \zeta_0)$

Random variable	**Asymptotic variance**
$n(\hat{\zeta}_n - \zeta_0)$	7.62
$n(\zeta_n^* - \zeta_0)$	63.98

LEMMA 4.3. *Unconditionally*:

(i) $\tilde{E}_n \rightsquigarrow \tilde{E}^*$ *in* $\mathcal{D}_K$ *for any compact rectangle* $K \subset \mathbb{R}^3$;
(ii) $\tilde{\Delta}_n^* = \operatorname{sargmax}_{h\in\mathbb{R}^3}\{\tilde{E}_n(h)\} \rightsquigarrow \operatorname{sargmax}_{h\in\mathbb{R}^3}\{\tilde{E}^*(h)\}$.

As a consequence, if the ECDF bootstrap is consistent, the variance of $\operatorname{sargmax}_{h\in\mathbb{R}^3}\{\tilde{E}^*(h)\}$ *must be twice that of* $\operatorname{sargmax}_{h\in\mathbb{R}^3}\{E^*(h)\}$.

As analytic expressions for the asymptotic variances of $n(\zeta_n^* - \zeta_0)$ and $n(\hat{\zeta}_n - \zeta_0)$ are not known, we use simulations to compute them. As an illustration, we take $\varepsilon \sim N(0, 1)$, $Z \sim N(0, 1)$, $\alpha_0 = -1$, $\beta_0 = 1$ and $\zeta_0 = 0$ in (1). We approximate the limiting variances with the sample variances computed from 20,000 observations from each of the two asymptotic distributions. Our results are summarized in Table 1, which immediately shows that the asymptotic variance of $n(\zeta_n^* - \zeta_0)$ is not twice that of $n(\hat{\zeta}_n - \zeta_0)$. Thus the ECDF bootstrap cannot be consistent.

4.2. *Scheme* 2 (*Bootstrapping "residuals"*). Another resampling procedure that arises naturally in a regression setup is bootstrapping "residuals." As with scheme 1, bootstrapping the "residuals" fixing the covariates is also *inconsistent*. Heuristically speaking, the resampling distribution fails to approximate the density of the predictor at the change-point ζ_0 at rate-n, and this leads to the inconsistency.

Recall the notation of Section 2. The following lemma (proved in Section A.2.12 of [25]) will be useful in the analysis of the smoothed bootstrap procedure.

LEMMA 4.4. *Let G and φ be, respectively, the distribution and characteristic functions of ε. Then*:

(i) *for any $\eta > 0$ we have that* $\sup_{|\xi|\le\eta}\{|\int e^{i\xi x}\, d\mathbb{P}_n^\varepsilon(x) - \varphi(\xi)|\} \xrightarrow{a.s.} 0$;
(ii) $\|\mathbb{P}_n^\varepsilon - G\|_{\mathbb{R}} \xrightarrow{a.s.} 0$;
(iii) $\int x^2\, d\mathbb{P}_n^\varepsilon(x) \xrightarrow{a.s.} \sigma^2$;
(iv) $\int |x|\, d\mathbb{P}_n^\varepsilon(x) \xrightarrow{a.s.} \mathbb{P}(|\varepsilon|)$;
(v) *if* $\mathbb{P}(|\varepsilon|^3) < \infty$, *then* $\overline{\lim}_{n\to\infty} \int |x|^3\, d\mathbb{P}_n^\varepsilon(x) < \infty$ *almost surely.*

The next result (proved in Section A.2.13 of [25]) shows that the bootstrapped least squares estimators converge conditionally in probability with probability one.

PROPOSITION 4.2. *For scheme* 2, $\theta_n^* \xrightarrow[a.s.]{\mathbf{P}_{\mathfrak{X}}} \theta_0$.

Consider the following process:

$$\hat{E}_n(h) = -\sum_{j=1}^{n}\left\{Y_{n,j}^* - \left(\hat{\alpha}_n + \frac{h_1}{\sqrt{n}}\right)\mathbf{1}_{Z_j \le \hat{\zeta}_n + h_3/n} - \left(\hat{\beta}_n + \frac{h_2}{\sqrt{n}}\right)\mathbf{1}_{Z_j > \hat{\zeta}_n + h_3/n}\right\}^2 + \sum_{j=1}^{n}(\varepsilon_{n,j}^*)^2.$$

Then for n large enough we have that

$$\big(\sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n), n(\zeta_n^* - \hat{\zeta}_n)\big) = \operatorname*{sargmax}_{h \in \mathbb{R}^3}\{\hat{E}_n(h)\}.$$

Next we argue that the sequence $(\hat{E}_n)_{n=1}^{\infty}$ does not have a weak limit in probability, and therefore distributional convergence of their corresponding smallest minimizers seems unreasonable.

LEMMA 4.5. *There is a compact rectangle $K \subset \mathbb{R}^3$ such that, conditional on the data, the sequence of processes $(\hat{E}_n)_{n=1}^{\infty}$ does not have any weak limit in probability in $\mathcal{D}_K$.*

PROOF. The proof is analogous to the proof of Lemma 4.2. We again consider the number $h_* > 0$ defined in the statement of Lemma 4.1(ii) and take $K \subset \mathbb{R}^3$ to be any compact rectangle containing the point $(0, 0, h_*)$. To prove the theorem it suffices to show that, conditional on the data, the sequence $(\hat{E}_n(0, 0, h_3))_{n=1}^{\infty}$ does not have a weak limit in probability whenever $h_3 \ge h_*$ and $(0, 0, h_3) \in K$. The (conditional) characteristic function of $\hat{E}_n(0, 0, h_3)$ is given by

$$\left(\int e^{i2(\hat{\alpha}_n - \hat{\beta}_n)\xi x - i\xi(\hat{\alpha}_n - \hat{\beta}_n)^2}\, d\mathbb{P}_n^{\varepsilon}(x)\right)^{n\mathbb{P}_n(\hat{\zeta}_n < Z \le \hat{\zeta}_n + h_3/n)}. \tag{11}$$

Now, Lemma 4.4 and the strong consistency of the least squares estimator imply that

$$\int e^{i2(\hat{\alpha}_n - \hat{\beta}_n)\xi x - i\xi(\hat{\alpha}_n - \hat{\beta}_n)^2}\, d\mathbb{P}_n^{\varepsilon}(x) \xrightarrow{\text{a.s.}} e^{-i\xi(\alpha_0 - \beta_0)^2}\varphi\big(2(\alpha_0 - \beta_0)\xi\big),$$

where φ is the characteristic function of ε. Thus, for $\xi \ne 0$ in a neighborhood of the origin, (11) will converge iff $n\mathbb{P}_n(\hat{\zeta}_n < Z \le \hat{\zeta}_n + \frac{h_3}{n})$ converges. But, from Lemma 4.1(ii), we know that this is not the case. $\square$

5. Consistent bootstrap procedures. Here we will prove that the "smoothed bootstrap" (scheme 3) and the m out of n bootstrap (scheme 4) procedures yield valid methods for constructing confidence intervals for θ_0.

5.1. *Scheme* 3 (*smoothed bootstrap*). To show that scheme 3 achieves consistency we appeal to Propositions 3.1, 3.2 and 3.3 by proving that the regularity conditions (I)–(VIII) of Section 3 hold for this scheme. Recall the description of this bootstrap procedure given in Section 2. Let $\hat{f}_n$ and $\hat{F}_n$ be the estimated smoothed density and distribution function of Z, respectively. For $I := [c, d] \subset \mathbb{R}$, a compact interval such that $\zeta_0 \in (c, d)$, we require the following two properties of $\hat{f}_n$ and $\hat{F}_n$:

$$\|\hat{F}_n - F\|_{\mathbb{R}} \xrightarrow{\text{a.s.}} 0; \tag{12}$$

$$\|\hat{f}_n - f\|_I \xrightarrow{\text{a.s.}} 0. \tag{13}$$

We would want to highlight that these conditions are fulfilled by many density estimation procedures. In particular, they hold when the density f is continuous, and we let $\hat{f}_n$ be the kernel density estimator constructed from a suitable choice of kernel and bandwidth (e.g., see [28]).

Let $\theta_n = \hat{\theta}_n$, $m_n = n$ and $\mathbb{Q}_n$ be the distribution that generates the bootstrap sample. Observe that under $\mathbb{Q}_n$, $\tilde{\varepsilon}_n$ and Z are independent and that Z is a continuous random variable with density $\hat{f}_n$. The next two results show that the bootstrapped least squares estimator achieves the right rate of convergence and has the right asymptotic distribution.

PROPOSITION 5.1. *If* (12) *and* (13) *hold*, *then w.p.* 1, *the sequence of conditional distributions of* $(\sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n), n(\zeta_n^* - \hat{\zeta}_n))$, *given the data*, *is tight*.

PROOF. We will show that conditions (I)–(V) in Section 3 hold w.p. 1 for the bootstrap measures arising in this scheme. Note that (IV) is a consequence of the almost sure convergence of the least squares estimators. That $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \xrightarrow{\text{a.s.}} 0$ follows immediately from the fact that $\|\hat{F}_n - F\|_\infty \xrightarrow{\text{a.s.}} 0$. Now, for any $g = y\psi \in \mathcal{G}$ with $\psi \in \mathcal{F}$, we have

$$\mathbb{Q}_n(g) = \hat{\alpha}_n \mathbb{Q}_n(\mathbf{1}_{Z \le \hat{\zeta}_n}\psi) + \hat{\beta}_n \mathbb{Q}_n(\mathbf{1}_{Z > \hat{\zeta}_n}\psi),$$
$$\mathbb{P}(g) = \alpha_0 \mathbb{P}(\mathbf{1}_{Z \le \zeta_0}\psi) + \beta_0 \mathbb{P}(\mathbf{1}_{Z > \zeta_0}\psi),$$

from which we see that

$$\begin{aligned}\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{G}} \le{}& (|\hat{\alpha}_n - \alpha_0| + |\hat{\beta}_n - \beta_0|) + (|\alpha_0| + |\beta_0|)\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \\ &+ (|\alpha_0| + |\beta_0|) \int_{\mathbb{R}} |\mathbf{1}_{z \le \hat{\zeta}_n} - \mathbf{1}_{z \le \zeta_0}| \hat{f}_n(z)\, dz.\end{aligned}$$

Lebesgue's dominated convergence theorem shows that the last integral goes almost surely to zero and the strong consistency of the least squares estimators and property (I) now yields $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{G}} \xrightarrow{\text{a.s.}} 0$. Finally, we can write any $h \in \mathcal{H}$ in the form $h = y^2\psi$ for some $\psi \in \mathcal{F}$. Using this representation we obtain

$$\mathbb{Q}_n(h) = \hat{\alpha}_n^2 \mathbb{Q}_n(\mathbf{1}_{Z \le \hat{\zeta}_n}\psi) + \hat{\beta}_n^2 \mathbb{Q}_n(\mathbf{1}_{Z > \hat{\zeta}_n}\psi) + \mathbb{P}_n^{\varepsilon}(\tilde{\varepsilon}_n^2)\mathbb{Q}_n(\psi),$$
$$\mathbb{P}(h) = \alpha_0^2 \mathbb{P}(\mathbf{1}_{Z \le \zeta_0}\psi) + \beta_0^2 \mathbb{P}(\mathbf{1}_{Z > \zeta_0}\psi) + \sigma^2 \mathbb{P}(\psi),$$

and the triangle inequality then implies that

$$\begin{aligned}
&\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{H}} \\
&\quad \le (|\hat{\alpha}_n^2 - \alpha_0^2| + |\hat{\beta}_n^2 - \beta_0^2|) + (\alpha_0^2 + \beta_0^2 + \sigma^2)\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \\
&\qquad + |\mathbb{P}_n^{\varepsilon}(\tilde{\varepsilon}_n^2) - \mathbb{P}(\varepsilon^2)| + (\alpha_0^2 + \beta_0^2) \int_{\mathbb{R}} |\mathbf{1}_{z \le \hat{\zeta}_n} - \mathbf{1}_{z \le \zeta_0}| \hat{f}_n(z)\, dz \\
&\quad \xrightarrow{\text{a.s.}} 0.
\end{aligned}$$

It remains to show (V). Observe that (4) and (5) hold automatically because under $\mathbb{Q}_n$, $\tilde{\varepsilon}_n$ and Z are independent. Hence, we are only required to show that (3) holds w.p. 1. As (13) holds, we have $\inf_{\zeta \in [c,d]}\{\hat{f}_n(\zeta)\} \xrightarrow{\text{a.s.}} \inf_{\zeta \in [c,d]}\{f(\zeta)\} > 0$. The mean value theorem implies that for any $\zeta, \xi \in [c, d]$, there is $\vartheta \in [0, 1]$ such that $|\hat{F}_n(\zeta) - \hat{F}_n(\xi)| = |\xi - \zeta| \hat{f}_n(\zeta + \vartheta(\xi - \zeta))$. It follows that for $\eta > 0$ small enough,

$$\inf_{0 < |\zeta - \hat{\zeta}_n| < \delta^2} \left\{ \frac{1}{|\zeta - \hat{\zeta}_n|} |\hat{F}_n(\zeta) - \hat{F}_n(\hat{\zeta}_n)| \right\} \ge \inf_{\zeta \in [c,d]} \{\hat{f}_n(\zeta)\} \qquad \text{for all } n \in \mathbb{N},$$

and consequently (V) holds w.p. 1 for all $\delta < \eta$ for all large n. □

PROPOSITION 5.2. *For scheme* 3, *provided that* (12) *and* (13) *hold*, *conditions* (I)–(VIII) *are satisfied with probability one*, *and thus*

$$\big(\sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n), n(\zeta_n^* - \hat{\zeta}_n)\big)' \rightsquigarrow \operatorname*{sargmax}_{h \in \mathbb{R}^3}\{E^*(h)\} \qquad \textit{almost surely.}$$

PROOF. We already know that conditions (I)–(V) hold w.p. 1. Condition (VII) holds automatically because Z and $\tilde{\varepsilon}_n$ are independent under $\mathbb{Q}_n$ and $\mathbb{Q}_n(\tilde{\varepsilon}_n) = 0$. Lemma 4.4(v) implies that condition (VIII) holds a.s. It remains to prove (VI).

Write $I = [c, d]$, and consider the sequence of events $\{A_N\}_{N \in \mathbb{N}}$ given by

$$\begin{aligned}
A_N = &\left[\hat{\zeta}_n - \frac{\delta}{n}, \hat{\zeta}_n + \frac{\eta}{n} \in I, \text{ almost always, for all } \delta, \eta \in (0, N)\right] \\
&\cap [\|\hat{f}_n - f\|_I \to 0].
\end{aligned}$$

Fix $N \in \mathbb{N}$, let ψ be the function $\psi(x) = e^{i\xi x}$ for some $\xi \in \mathbb{R}$ or the function $\psi(x) = |x|^p$, $p = 1, 2$, and $\eta, \delta > 0$ be any positive real numbers smaller than N.

Then

$$m_n \mathbb{Q}_n(\psi(\tilde{\varepsilon}_n)\mathbf{1}_{\zeta_n-\delta/n<Z\le\zeta_n+\eta/n}) = n\mathbb{P}_n^{\varepsilon}(\psi)\int_{\hat{\zeta}_n-\delta/n}^{\hat{\zeta}_n+\eta/n} \hat{f}_n(x)\,dx.$$

Lemma 4.4 implies that $\mathbb{P}_n^{\varepsilon}(\psi) \xrightarrow{\text{a.s.}} \mathbb{P}(\psi(\varepsilon))$. And, when A_N holds, we also have

$$n\left|\int_{\hat{\zeta}_n-\delta/n}^{\hat{\zeta}_n+\eta/n} \hat{f}_n(x)\,dx - \int_{\hat{\zeta}_n-\delta/n}^{\hat{\zeta}_n+\eta/n} f(x)\,dx\right| \le 2N\|\hat{f}_n - f\|_I \to 0.$$

Hence, condition (VI) holds for all $0<\delta, \eta<N$ on A_N. But the strong consistency of the least squares estimators and the conditions on $\hat{f}_n$ imply that each of these events have probability one. Therefore, $\mathbf{P}(\bigcap_{N\in\mathbb{N}} A_N) = 1$. Hence, condition (VI) holds w.p. 1, and the result follows from an application of Proposition 3.3. □

5.2. *Scheme* 4 (*m out of n bootstrap*). For this scheme we will again use the framework established in Section 3. We take $(m_n)_{n=1}^{\infty}$ to be any sequence of natural numbers which increases to infinity, $\hat{\theta}_n = \theta_n$ and $\mathbb{Q}_n = \mathbb{P}_n$. The next result (proved in Section A.2.17 of [25]) shows the *weak consistency* of this procedure.

PROPOSITION 5.3. *If $m_n = o(n)$ and $m_n \uparrow \infty$, then conditions* (I)–(VIII) *hold* (*in probability*), *and we have*

$$\begin{aligned}&\big(\sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n), n(\zeta_n^* - \hat{\zeta}_n)\big)'\\ &\qquad \rightsquigarrow \operatorname*{sargmax}_{h\in\mathbb{R}^3}\{E^*(h)\} \qquad \textit{in probability.}\end{aligned}$$

REMARK. To prove Proposition 5.3, we, in fact, show that for every subsequence $(n_k)_{k=1}^{\infty}$, there is a further subsequence $(n_{k_s})_{s=1}^{\infty}$, such that (I)–(VIII) hold w.p. 1 for $(n_{k_s})_{s=1}^{\infty}$, and the above result holds almost surely along the subsequence $(n_{k_s})_{s=1}^{\infty}$.

6. Simulation experiments. In this section we report the finite sample performance of the different bootstrap schemes on simulated data. We simulated random draws from four different models following (1). Each of these corresponded to choosing different pairs (F, G) of distributions for Z and ε (having mean 0), respectively. The pairs considered were $(N(0,2), N(0,1))$, $(4B(4,6)-2, N(0,1))$, $(4B(4,6)-2, \mathrm{Unif}(-1,1))$ and $(4B(4,6)-2, \Gamma(4,2)-2)$, where $B(\cdot,\cdot)$ and $\Gamma(\cdot,\cdot)$ denote the beta and gamma distributions, respectively. For all the simulations we considered $\theta_0 = (\alpha_0, \beta_0, \zeta_0) = (-1, 1, 0)$.

For each of these models, we considered 1,000 random samples of sizes $n = 50, 200, 500$. For each sample, and for each of the bootstrap schemes, we took $4n$ bootstrap replicates to approximate the bootstrap distribution. Table 2 provides the estimated coverage proportions and average lengths of nominal 95% CIs obtained using the 4 different bootstrap schemes for each of the four models.

TABLE 2
The estimated coverage probabilities and average lengths of nominal 95% CIs for ζ_0 obtained using the four different bootstrap schemes for each of the four models

	$n = 50$		$n = 200$		$n = 500$	
Scheme	**Coverage**	**Avg length**	**Coverage**	**Avg length**	**Coverage**	**Avg length**
			$Z \sim N(0, 2), \varepsilon \sim N(0, 1)$			
ECDF	0.83	1.14	0.79	0.22	0.81	0.08
Smoothed	0.94	0.94	0.95	0.19	0.95	0.07
FDR	0.83	0.76	0.86	0.16	0.90	0.06
$\lceil n^{4/5} \rceil$	0.87	0.87	0.91	0.23	0.91	0.08
$\lceil n^{9/10} \rceil$	0.85	1.02	0.87	0.21	0.87	0.079
$\lceil n^{14/15} \rceil$	0.85	1.05	0.84	0.21	0.86	0.08
			$Z \sim 4B(4, 6) - 2, \varepsilon \sim N(0, 1)$			
ECDF	0.80	0.54	0.80	0.11	0.81	0.04
Smoothed	0.96	0.46	0.94	0.11	0.95	0.47
FDR	0.73	0.32	0.77	0.08	0.79	0.03
$\lceil n^{4/5} \rceil$	0.88	0.53	0.89	0.11	0.90	0.04
$\lceil n^{9/10} \rceil$	0.85	0.54	0.86	0.11	0.88	0.04
$\lceil n^{14/15} \rceil$	0.83	0.55	0.84	0.11	0.87	0.04
			$Z \sim 4B(4, 6) - 2, \varepsilon \sim \text{Unif}(-1, 1)$			
ECDF	0.80	0.40	0.80	0.08	0.81	0.03
Smoothed	0.94	0.33	0.95	0.08	0.96	0.04
FDR	0.75	0.26	0.77	0.06	0.81	0.02
$\lceil n^{4/5} \rceil$	0.88	0.36	0.88	0.09	0.91	0.04
$\lceil n^{9/10} \rceil$	0.85	0.39	0.85	0.08	0.87	0.03
$\lceil n^{14/15} \rceil$	0.83	0.39	0.84	0.08	0.85	0.03
			$Z \sim 4B(4, 6) - 2, \varepsilon \sim \Gamma(4, 2) - 2$			
ECDF	0.80	0.49	0.80	0.09	0.81	0.04
Smoothed	0.93	0.36	0.95	0.08	0.96	0.03
FDR	0.76	0.30	0.77	0.06	0.80	0.02
$\lceil n^{4/5} \rceil$	0.87	0.43	0.88	0.10	0.91	0.03
$\lceil n^{9/10} \rceil$	0.85	0.46	0.84	0.09	0.88	0.03
$\lceil n^{14/15} \rceil$	0.83	0.48	0.85	0.09	0.85	0.03

At this point, we want to make some remarks about the computation of the estimators. We used a kernel density estimator based on the Gaussian kernel and chose the bandwidth by the so-called "normal reference rule" (see [24], page 131). In the case of the m out of n bootstrap, we did not use any data driven choice of m_n, but tried 3 different possibilities: $\lceil n^{4/5} \rceil$, $\lceil n^{9/10} \rceil$ and $\lceil n^{14/15} \rceil$. We will refer to the fixed-design bootstrapping of residuals scheme by FDR.

We can see from the table that the smoothed bootstrap scheme outperforms all the others in terms of coverage. It must also be noted that this is achieved without a relative increase in the lengths of the intervals. The m out of n bootstrap with $\lceil n^{4/5} \rceil$ also performs reasonably well. It clearly outperforms all other m out of n schemes as well as ECDF and FDR bootstrap procedures (which are inconsistent). Table 2 clearly shows the implications of using the two most common bootstrap procedures—the estimated coverage probabilities of the nominal 95% CIs constructed from the ECDF and the FDR bootstrap schemes suffer from drastic under-coverage (varying between 0.75 to 0.85).

Figure 1 shows the histograms of the distribution of $n(\hat{\zeta}_n - \zeta_0)$ (obtained from 1,000 random samples) and its bootstrap estimates obtained from the 4 different bootstrap schemes (using 2,000 bootstrap samples each) from a single data set of size $n = 500$ drawn randomly from model (1) with $Z \sim 4B(4, 6) - 2$, $\varepsilon \sim \Gamma(4, 2) - 2$. The histograms clearly show that the smoothed bootstrap (top right panel) provides, by far, the best approximation to both, the actual (top middle panel) and the limiting distributions (top left panel). In fact, the histograms of the distribution of $n(\hat{\zeta}_n - \zeta_0)$ and the corresponding smoothed bootstrap estimate are almost indistinguishable. The m out of n approach, although guaranteed to converge, lacks the efficiency of the smoothed bootstrap. This may be due to the fact that we do not have an optimal way of choosing the tuning parameter m, the block

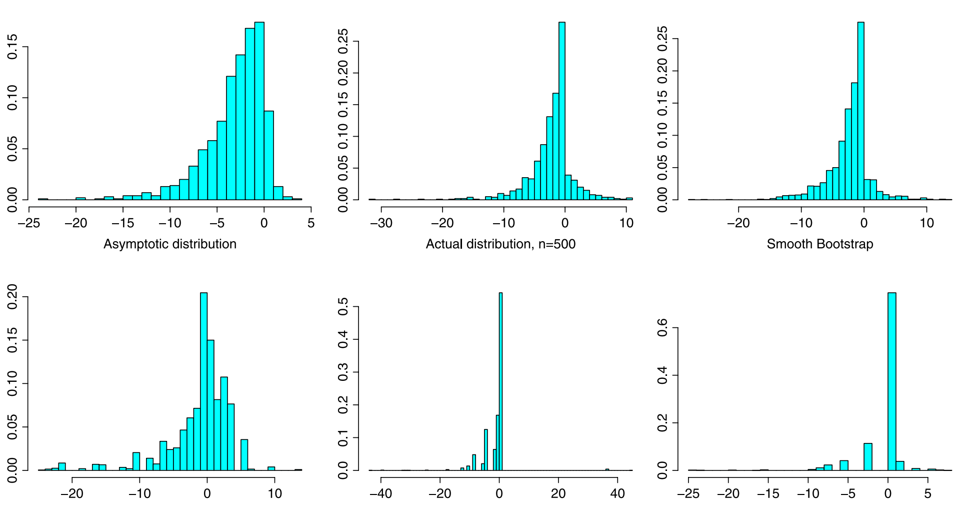

FIG. 1. *Histograms of the distribution of* $n(\hat{\zeta}_n - \zeta_0)$ *and its bootstrap estimates: the asymptotic distribution of* $n(\hat{\zeta}_n - \zeta_0)$ (*top left*); *the actual distribution of* $n(\hat{\zeta}_n - \zeta_0)$ (*top middle*); *the distribution of* $n(\zeta_n^* - \hat{\zeta}_n)$ *for the smoothed* (*top right*); *ECDF* (*bottom middle*) *and FDR* (*bottom right*) *schemes; the distribution of* $m_n(\zeta_n^* - \hat{\zeta}_n)$, $m_n = \lceil n^{4/5} \rceil$ (*bottom left*).

size. The smoothed bootstrap, although it requires the choice of the smoothing bandwidth, is much more robust against different choices of the tuning parameter and has a clear advantage over the m out of n bootstrap procedure.

7. More general change-point regression models. In this section we mention some of the broader implications of our analysis of (1) in the context of more general change-point models in regression. We can consider a model of the form

$$Y = \psi_{\alpha_0}(W, Z)\mathbf{1}_{Z\le\zeta_0} + \xi_{\beta_0}(W, Z)\mathbf{1}_{Z>\zeta_0} + \varepsilon, \tag{14}$$

where Z is a continuous random variable; W is a random vector of covariates independent of Z; $\alpha_0 \in \mathbb{R}^p$ and $\beta_0 \in \mathbb{R}^q$ are two unknown Euclidian parameters; $\psi_\alpha(w, z)$ and $\xi_\beta(w, z)$ are known real-valued functions continuous in (w, z) and twice continuously differentiable in α and β, respectively; $\zeta_0 \in [a, b] \subset \mathrm{supp}(Z) \subset \mathbb{R}$ is the change-point; ε is a continuous random variable, independent of (W, Z), with zero expectation and finite variance $\sigma^2 > 0$. We also assume that W and Z are independent. We assume that $\psi_{\alpha_0}(W, Z)$ is identifiable from $\xi_{\beta_0}(W, Z)$ and that the least squares problems $\min_{\alpha\in\mathbb{R}^p} \sum_{Z_j\le\zeta}(Y_j - \psi_\alpha(W_j, Z_j))^2$ and $\min_{\beta\in\mathbb{R}^q} \sum_{Z_j>\zeta}(Y_j - \xi_\beta(W_j, Z_j))^2$ are well posed for every possible data set $\{(Y_1, Z_1, W_1), \dots, (Y_n, Z_n, W_n)\}$ and any $\zeta \in \mathrm{supp}(Z)^\circ$. We also assume that $\psi_{\alpha_0}(w, \zeta_0) \ne \xi_{\beta_0}(w, \zeta_0)$ for every value of w.

Like in the simple case, the method of least squares can be used to compute estimators $\hat\alpha_n$, $\hat\beta_n$ and $\hat\zeta_n$. One simply takes the minimizer $(\hat\alpha_n, \hat\beta_n, \hat\zeta_n)$ of

$$\sum_{j=1}^n \bigl(Y_j - \psi_\alpha(W_j, Z_j)\mathbf{1}_{Z_j\le\zeta} + \xi_\beta(W_j, Z_j)\mathbf{1}_{Z_j>\zeta}\bigr)^2$$

with the smallest ζ-component.

Since the simple model (1) is a particular case of (14), one can immediately conclude from our analysis that the usual ECDF and residual bootstrap procedures will not be consistent. However, the smoothed bootstrap can be adapted to produce valid CIs. The modified scheme can be described as follows:

(1) Choose some procedure (e.g., kernel density estimation) to build a distribution $\hat F_n$ with density $\hat f_n$ such that $\hat f_n \to f$ uniformly on compact intervals w.p. 1, where f is the density of Z. Let $\mathbb{P}_n^\varepsilon$ and $\mathbb{P}_n^W$ be the empirical measures of the centered residuals (as in the description of scheme 2 in Section 2) and $W_1, \dots, W_n$, respectively.

(2) Get i.i.d. replicates $Z^*_{n,1}, \dots, Z^*_{n,n}$ from $\hat F_n$ and sample, independently, $\varepsilon^*_{n,1}, \dots, \varepsilon^*_{n,n} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_n^\varepsilon$ and $W^*_{n,1}, \dots, W^*_{n,n} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_n^W$. We could also keep the W_i's fixed, that is, $W^*_{n,i} = W_i$.

(3) Define $Y^*_{n,j} := \psi_{\hat\alpha_n}(W^*_{n,j}, Z^*_{n,j})\mathbf{1}_{Z^*_{n,j}\le\hat\zeta_n} + \xi_{\hat\beta_n}(W^*_{n,j}, Z^*_{n,j})\mathbf{1}_{Z^*_{n,j}>\hat\zeta_n} + \varepsilon^*_{n,j}$ for all $j = 1, \dots, n$.

(4) Compute the bootstrap least squares estimator $(\alpha_n^*, \beta_n^*, \zeta_n^*)$ by taking the minimizer of $\sum_{j=1}^n (Y_{n,j}^* - \psi_\alpha(W_{n,j}^*, Z_{n,j}^*)\mathbf{1}_{Z_{n,j}^* \le \zeta} - \xi_\beta(W_{n,j}^*, Z_{n,j}^*)\mathbf{1}_{Z_{n,j}^* > \zeta})^2$ with the smallest ζ-component.

(5) Approximate the distribution of $n(\hat{\zeta}_n - \zeta_0)$ with the (conditional) distribution of $n(\zeta_n^* - \hat{\zeta}_n)$.

Although our analysis indicates that this smoothed bootstrap procedure will be consistent, it is difficult to use our methods to prove consistency in such generality. However, the proof for the simple model (1) can be adapted to cover the case of parametric additive models, that is, when $\psi_\alpha(w, z)$ and $\xi_\beta(w, z)$ are of the form $\psi_\alpha(w, z) = \sum_{j=1}^p \alpha_j g_j(w, z)$ and $\xi_\beta(w, z) = \sum_{k=1}^q \beta_k h_k(w, z)$, where g_j, h_k, $j = 1, \dots, p$, $k = 1, \dots, q$, are known smooth functions.

APPENDIX

A.1. Proof of Proposition 3.3. We express the process $\hat{E}_n$ as the sum of the four terms $\hat{A}_n$, $\hat{B}_n$, $\hat{C}_n$ and $\hat{D}_n$ where

$$
\begin{aligned}
\hat{A}_n(h_1, h_3) &:= 2h_1\sqrt{m_n}\mathbb{P}_n^*\big(\tilde{\varepsilon}_n \mathbf{1}_{Z \le \zeta_n \wedge (\zeta_n + h_3/m_n)}\big) - h_1^2 \mathbb{P}_n^*\big(\mathbf{1}_{Z \le \zeta_n \wedge (\zeta_n + h_3/m_n)}\big), \\
\hat{B}_n(h_2, h_3) &:= 2h_2\sqrt{m_n}\mathbb{P}_n^*\big(\tilde{\varepsilon}_n \mathbf{1}_{Z > \zeta_n \vee (\zeta_n + h_3/m_n)}\big) - h_2^2 \mathbb{P}_n^*\big(\mathbf{1}_{Z > \zeta_n \vee (\zeta_n + h_3/m_n)}\big), \\
\hat{C}_n(h_2, h_3) &:= -2m_n\bigg(\alpha_n - \beta_n + \frac{h_2}{\sqrt{m_n}}\bigg)\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{\zeta_n + h_3/m_n < Z \le \zeta_n}) \\
&\qquad - m_n\bigg(\alpha_n - \beta_n + \frac{h_2}{\sqrt{m_n}}\bigg)^2 \mathbb{P}_n^*(\mathbf{1}_{\zeta_n + h_3/m_n < Z \le \zeta_n}), \\
\hat{D}_n(h_1, h_3) &:= -2m_n\bigg(\beta_n - \alpha_n + \frac{h_1}{\sqrt{m_n}}\bigg)\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{\zeta_n < Z \le \zeta_n + h_3/m_n}) \\
&\qquad - m_n\bigg(\beta_n - \alpha_n + \frac{h_1}{\sqrt{m_n}}\bigg)^2 \mathbb{P}_n^*(\mathbf{1}_{\zeta_n < Z \le \zeta_n + h_3/m_n}).
\end{aligned} \tag{15}
$$

We define another process $E_n^* := A_n^* + B_n^* + C_n^* + D_n^*$ where

$$
\begin{aligned}
A_n^*(h_1) &:= 2h_1\sqrt{m_n}\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{Z \le \zeta_n}) - h_1^2 \mathbb{P}_n^*(\mathbf{1}_{Z \le \zeta_n}), \\
B_n^*(h_2) &:= 2h_2\sqrt{m_n}\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{Z > \zeta_n}) - h_2^2 \mathbb{P}_n^*(\mathbf{1}_{Z > \zeta_n}), \\
C_n^*(h_3) &:= -2m_n(\alpha_n - \beta_n)\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{\zeta_n + h_3/m_n < Z \le \zeta_n}) \\
&\qquad - m_n(\alpha_n - \beta_n)^2 \mathbb{P}_n^*(\mathbf{1}_{\zeta_n + h_3/m_n < Z \le \zeta_n}), \\
D_n^*(h_3) &:= -2m_n(\beta_n - \alpha_n)\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{\zeta_n < Z \le \zeta_n + h_3/m_n}) \\
&\qquad - m_n(\beta_n - \alpha_n)^2 \mathbb{P}_n^*(\mathbf{1}_{\zeta_n < Z \le \zeta_n + h_3/m_n}).
\end{aligned}
$$

We work with E_n^* instead of $\hat{E}_n$ as their difference approaches uniformly to 0 in probability, as shown in the next lemma (proved in Section A.2.3 of [25]), and the asymptotic distribution of E_n^* is easier to derive.

LEMMA A.1. *Let $K \subset \mathbb{R}^3$ be a compact rectangle. If conditions* (I)–(IV) *and* (VI) *hold, then* $\|E_n^* - \hat{E}_n\|_K \xrightarrow{\mathbf{P}} 0$. *Therefore,* $E_n^* - \hat{E}_n \xrightarrow{\mathbf{P}} 0$ *as random elements of $\mathcal{D}_K$.*

As a first step to finding the asymptotic distribution of $(E_n^*)_{n=1}^\infty$, we show that the random sequence is tight in the Skorohod space $\mathcal{D}_K$ for any compact rectangle $K \subset \mathbb{R}^3$. Let $\tilde{\mathcal{D}}_I$, $I \subset \mathbb{R}$, denote the space of real valued càdlàg functions on I. The proof of the next result is given in Section A.2.4. of [25].

LEMMA A.2. *Let $I \subset \mathbb{R}$ be a compact interval and assume that conditions* (I)–(VIII) *hold. Then, the sequence of $\mathbb{R}^6$-valued processes*

$$\begin{aligned} \Xi_n(t) := \big(&\sqrt{m_n}\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{Z \le \zeta_n}), \sqrt{m_n}\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{Z > \zeta_n}), \\ &m_n\mathbb{P}_n^*(\mathbf{1}_{\zeta_n + t/m_n < Z \le \zeta_n}), m_n\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{\zeta_n + t/m_n < Z \le \zeta_n}), \\ &m_n\mathbb{P}_n^*(\mathbf{1}_{\zeta_n < Z \le \zeta_n + t/m_n}), m_n\mathbb{P}_n^*(\tilde{\varepsilon}_n \mathbf{1}_{\zeta_n < Z \le \zeta_n + t/m_n})\big)' \end{aligned} \tag{16}$$

is uniformly tight in $\mathbb{R}^2 \times \tilde{\mathcal{D}}_I^4$. Also, if $K \subset \mathbb{R}^3$ is a compact rectangle, the sequence $(E_n^*)_{n=1}^\infty$ is uniformly tight in $\mathcal{D}_K$.

It now suffices to show convergence of the finite-dimensional distributions of the processes E_n^* to the finite-dimensional distributions of $E^* \in \mathcal{D}_K$ to conclude that E_n^* converges weakly to E^* (and thus $\hat{E}_n$ too). Then an application of the continuous mapping theorem for the smallest argmax functional (see Lemma A.3 of [25]) gives the weak convergence of $h_n^* := \operatorname{sargmax} \hat{E}_n(h)$. The application of the lemma requires the weak convergence of processes $(\hat{E}_n)_{n=1}^\infty$ to E^* and also the weak convergence of their associated jump processes. Let $\mathcal{S}$ be the class of all piecewise constant, cádlág functions $\tilde{\psi} : \mathbb{R} \to \mathbb{R}$ that are continuous on the integers with $\tilde{\psi}(0) = 0$; $\tilde{\psi}$ has jumps of size 1, and $\tilde{\psi}(-t)$ and $\tilde{\psi}(t)$ are nondecreasing on $(0, \infty)$. For an interval I containing 0 in its interior, we write $\mathcal{S}_I = \{f|_I : f \in \mathcal{S}\}$. Define the $\mathcal{S}$-valued (pure jump) processes $\hat{J}_n$, J_n^* and J^* as

$$\begin{aligned} J_n^*(t) = \hat{J}_n(t) &:= m_n\mathbb{P}_n^*(\mathbf{1}_{\zeta_n + t/m_n < Z \le \zeta_n}) + m_n\mathbb{P}_n^*(\mathbf{1}_{\zeta_n < Z \le \zeta_n + t/m_n}), \\ J^*(t) &:= \nu_1(-t)\mathbf{1}_{t<0} + \nu_2(t)\mathbf{1}_{t \ge 0}. \end{aligned}$$

LEMMA A.3. *Let $I \subset \mathbb{R}$ be a compact interval and $K = A \times B \times I \subset \mathbb{R}^3$ a compact rectangle. If* (I)–(VIII) *hold, we have:*

1. $\Xi_n \rightsquigarrow \Xi$ *in* $\mathbb{R}^2 \times \tilde{\mathcal{D}}_I^4$;

2. $(E_n^*, J_n^*) \rightsquigarrow (E^*, J^*)$ *in* $\mathcal{D}_K \times \mathcal{S}_I$;
3. $(\hat{E}_n, \hat{J}_n) \rightsquigarrow (E^*, J^*)$ *in* $\mathcal{D}_K \times \mathcal{S}_I$.

For a proof of the convergence result, see Section A.2.5 of [25].

A.2. Proof of Lemma 4.1. We state two lemmas that will be crucial in the proof of this result. A proof of Lemma A.4 can be found in [25].

LEMMA A.4. *Let* $\lambda, B > 0$, $\rho \in (0, \frac{1}{2})$ *and* H_λ *be the distribution function of a Poisson random variable with mean* λ. *For each value of* λ *write* $L_{\lambda+B}^\rho = \min\{n \in \mathbb{N} : H_{\lambda+B}(n) > \rho\}$ *and* $U_\lambda^\rho = \max\{n \in \mathbb{N} : 1 - H_\lambda(n) > \rho\}$. *Then there is* $\lambda_* > 0$ *such that* $L_{\lambda+B}^\rho < U_\lambda^\rho$ *for all* $\lambda \geq \lambda_*$.

LEMMA A.5. *Let* $\lambda, B > 0$, $0 < \rho < \frac{1}{2}$, μ *and* ν *be two nondegenerate Borel probability measures on* $\mathbb{R}$ *and* $H_{\mu,\lambda}$ *denote the compound Poisson distribution with intensity* λ *and compounding distribution* μ. *For each value of* λ *write* $L_{\nu,\lambda+B}^\rho = \inf\{s \in \mathbb{R} : H_{\nu,\lambda+B}(s) \geq \rho\}$ *and* $U_{\mu,\lambda}^\rho = \sup\{s \in \mathbb{R} : 1 - H_{\mu,\lambda}(s) \geq \rho\}$. *In addition, assume that* $\int x^2 \nu(dx), \int x^2 \mu(dx) < \infty$ *and that* $\int x \nu(dx) \leq \int x \mu(dx)$. *Then there is* $\lambda_* > 0$ *such that* $L_{\nu,\lambda+B}^\rho < U_{\mu,\lambda}^\rho$ *for all* $\lambda \geq \lambda_*$. *Moreover, let* $0 < r < 1$, *and suppose that there is another Borel probability measure* γ *on* $\mathbb{R}$, *and define* $\nu_\gamma := \frac{rB}{\lambda+B}\gamma + \frac{\lambda+(1-r)B}{\lambda+B}\nu$ *and the corresponding constant* $L_{\nu_\gamma,\lambda+B}^\rho = \inf\{s \in \mathbb{R} : H_{\nu_\gamma,\lambda+B}(s) \geq \rho\}$. *Then there is* $\lambda_* > 0$ *such that* $L_{\nu_\gamma,\lambda+B}^\rho < U_{\mu,\lambda}^\rho$ *for all* $\lambda \geq \lambda_*$.

PROOF. Denote by Φ the standard normal distribution and $\mathbf{z}_\alpha$ the lower α-quantile of Φ [i.e., $\Phi(\mathbf{z}_\alpha) = \alpha$]. Also, write $c_\mu := \int x\mu(dx)$, $d_\mu := \int x^2 \mu(dx)$ and define the corresponding quantities c_ν and d_ν for ν. For any possible value of λ and μ denote by $T_{\mu,\lambda}$ a random variable with distribution $H_{\mu,\lambda}$. It is easily seen (as, e.g., in Theorem 2.1 of [20]) that $S_{\mu,\lambda} := \frac{T_{\mu,\lambda} - \lambda c_\mu}{\sqrt{\lambda d_\mu}} \rightsquigarrow \Phi$ as $\lambda \to \infty$. Since the standard normal distribution is continuous, the distributions of $S_{\mu,\lambda}$ converge uniformly on $\mathbb{R}$ to Φ as $\lambda \to \infty$.

Let $1 < \kappa < 1/(2\rho)$. Then, since the distributions of $S_{\mu,\lambda}$ converge uniformly to Φ, there is λ_1 such that $1 - \Phi(\frac{U_{\mu,\lambda}^\rho - \lambda c_\mu}{\sqrt{\lambda d_\mu}}) < \kappa\rho$ for $\lambda > \lambda_1$ and $\lambda_2 > 0$ such that $\Phi(\frac{L_{\nu,\lambda+B}^\rho - (\lambda+B)c_\nu}{\sqrt{(\lambda+B)d_\nu}}) < \kappa\rho$ for all $\lambda > \lambda_2$. These two inequalities in turn imply that

$$U_{\mu,\lambda}^\rho > \lambda c_\mu - \sqrt{\lambda d_\mu}\mathbf{z}_{\kappa\rho},$$
$$L_{\nu,\lambda+B}^\rho < (\lambda + B)c_\nu + \sqrt{(\lambda + B)d_\nu}\mathbf{z}_{\kappa\rho}.$$

Since $c_\mu \geq c_\nu$ we can find λ_3 such that

$$(\lambda + B)c_\nu + \sqrt{(\lambda + B)d_\nu}\mathbf{z}_{\kappa\rho} < \lambda c_\mu - \sqrt{\lambda d_\mu}\mathbf{z}_{\kappa\rho} \qquad \text{for all } \lambda \geq \lambda_3.$$

The first part of the result now follows by taking $\lambda_* := \lambda_1 \vee \lambda_2 \vee \lambda_3$. To prove the result for the measure ν_γ it suffices to see that we also have $\frac{T_{\nu_\gamma,\lambda+B}-(\lambda+B)c_{\nu_\gamma}}{\sqrt{(\lambda+B)d_{\nu_\gamma}}} \rightsquigarrow \Phi$, as $\lambda \to \infty$ (this is easily seen by analyzing the characteristic functions). The rest follows from the same argument used to prove the first part of the lemma. $\square$

PROOF OF LEMMA 4.1(i). Let $s < t$. Note that $(Z_n)_{n=1}^{\infty}$ is a collection of i.i.d. random variables and $n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \le \zeta_0 + \frac{t}{n})$ is permutation invariant, so the Hewitt–Savage 0–1 law (see page 496 of [3]) implies that any convergent subsequence must converge to a constant. On the other hand, using characteristic functions it can be shown that $n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \le \zeta_0 + \frac{t}{n}) \rightsquigarrow \mathrm{Poisson}((t-s)f(\zeta_0))$. Therefore, $(n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \le \zeta_0 + \frac{t}{n}))_{n=1}^{\infty}$ has no almost surely convergent subsequence.

PROOF OF (ii). Now, let $\delta \in (0, \frac{1}{4})$. From Proposition 3.2 we know that there is $B_\delta > 0$ such that $\mathbf{P}(n|\hat{\zeta}_n - \zeta_0| \le B_\delta) > 1 - \delta$ for any $n \in \mathbb{N}$. Choose $h > 2B_\delta$, and take any increasing sequence of natural numbers n_k. Write $\hat{T}_k = n_k\mathbb{P}_{n_k}(\hat{\zeta}_{n_k} < Z \le \hat{\zeta}_{n_k} + \frac{h}{n_k})$, $S_k = n_k\mathbb{P}_{n_k}(\zeta_0 - \frac{B_\delta}{n_k} < Z \le \zeta_0 + \frac{h+B_\delta}{n_k})$ and $T_k = n_k\mathbb{P}_{n_k}(\zeta_0 + \frac{B_\delta}{n_k} < Z \le \zeta_0 + \frac{h-B_\delta}{n_k})$. Then, $\{n_k|\hat{\zeta}_{n_k} - \zeta_0| \le B_\delta\} \subset \{S_k \ge \hat{T}_k \ge T_k\}$ and hence $\mathbf{P}(\hat{T}_k \ge T_k) \wedge \mathbf{P}(S_k \ge \hat{T}_k) > 1 - \delta$ for all k.

We know that $T_k \rightsquigarrow \mathrm{Poisson}((h - 2B_\delta)f(\zeta_0))$ and $S_k \rightsquigarrow \mathrm{Poisson}((h + 2B_\delta) \times f(\zeta_0))$, so in view of Lemma A.4 with $B = 4B_\delta f(\zeta_0)$ and $\lambda = (h - 2B_\delta)f(\zeta_0)$, there is a number $h_* > 2B_\delta$ large enough so that whenever $h \ge h_*$ we can find two numbers $N_{1,h} < N_{2,h} \in \mathbb{N}$ with the property that, $\underline{\lim}_{k\to\infty}\mathbf{P}(T_k > N_{2,h}) > 2\delta$ and $\underline{\lim}_{k\to\infty}\mathbf{P}(S_k \le N_{1,h}) > 2\delta$. Thus, for $h \ge h_*$, $\mathbf{P}(T_k > N_{2,h}) > 2\delta$ and $\mathbf{P}(S_k \le N_{1,h}) > 2\delta$ for all but a finite number of k's. Therefore, for any k large enough, $\mathbf{P}(T_k > N_{2,h}) \wedge \mathbf{P}(S_k \le N_{1,h}) > 2\delta$. Using the fact that $\mathbf{P}(S_k \ge \hat{T}_k \ge T_k) > 1 - \delta$ we get that $\mathbf{P}(\hat{T}_k \ge T_k > N_{2,h}) \wedge \mathbf{P}(N_{1,h} \ge S_k \ge \hat{T}_k) > \delta$ for all but finitely many k's. Thus, whenever $h \ge h_*$,

$$\mathbf{P}(\hat{T}_k \ge T_k > N_{2,h}, \text{i.o.}) \ge \delta \quad \text{and} \quad \mathbf{P}(N_{1,h} \ge S_k \ge \hat{T}_k, \text{i.o.}) \ge \delta.$$

But for every $k \in \mathbb{N}$, the events $\{\hat{T}_k \ge T_k > N_{2,h}\}$ and $\{N_{1,h} \ge S_k \ge \hat{T}_k\}$ are permutation-invariant on the i.i.d. random vectors $X_1, \dots, X_{n_k}$. Hence, the Hewitt–Savage 0–1 law implies that $\mathbf{P}(\hat{T}_k \ge T_k > N_{2,h}, \text{i.o.}) = 1$ and $\mathbf{P}(N_{1,h} \ge S_k \ge \hat{T}_k, \text{i.o.}) = 1$. Since $N_{1,h} < N_{2,h}$ it follows that $\hat{T}_k = n_k\mathbb{P}_{n_k}(\hat{\zeta}_{n_k} < Z \le \hat{\zeta}_{n_k} + h/n_k)$ does not have an almost sure limit. But the choice of the subsequence n_k was arbitrary and independent of h_*, so we can conclude that for any $h \ge h_*$, the sequence $\{n\mathbb{P}_n(\hat{\zeta}_n < Z \le \hat{\zeta}_n + \frac{h}{n})\}_{n=1}^{\infty}$ does not converge in probability. Proceeding analogously, we can prove the same for $\{n\mathbb{P}_n(\hat{\zeta}_n - \frac{h}{n} < Z \le \hat{\zeta}_n)\}_{n=1}^{\infty}$.

PROOF OF (iii). We introduce some notation, for any two Borel probability measures μ and ν on $\mathbb{R}$ we write $\mu \bigstar \nu$ for their convolution, and for $\lambda > 0$ we write CPoisson(μ, λ) for the compound Poisson distribution with intensity λ and compounding distribution μ. Let μ_α and μ_β be, respectively, the distributions under $\mathbb{P}$ of $\phi(\varepsilon + \alpha_0)$ and $\phi(\varepsilon + \beta_0)$.

Observe that depending on whether $t < 0$, $s < 0 < t$ or $s > 0$ we have that $n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\zeta_0+s/n<Z\le\zeta_0+t/n})$ converges weakly to CPoisson$(\mu_\alpha, (t-s)f(\zeta_0))$, CPoisson$(\mu_\alpha, sf(\zeta_0)) \bigstar$ CPoisson$(\mu_\beta, tf(\zeta_0))$ or CPoisson$(\mu_\beta, (t-s)f(\zeta_0))$, respectively. This follows easily from convergence of the corresponding characteristic functions. Considering that $\{(Y_n, Z_n)\}_{n=1}^\infty$ is a collection of i.i.d. random vectors and that $n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\zeta_0+s/n<Z\le\zeta_0+t/n})$ is permutation invariant for $(Y_1, Z_1), \dots, (Y_n, Z_n)$ the same argument as in (i) applies here as well.

PROOF OF (iv). The argument is quite similar to the one used to prove (ii). Assume without loss of generality that $\phi \le 0$. Let $\delta \in (0, \frac{1}{4})$ and $N \in \mathbb{N}$. From Proposition 3.2 we know that there is $B_\delta > 0$ such that $\mathbf{P}(n|\hat{\zeta}_n - \zeta_0| \le B_\delta) > 1 - \delta$ for any $n \in \mathbb{N}$. Choose $h > 2B_\delta$, and take any increasing sequence of natural numbers n_k. Write $\hat{T}_{k,h}^\phi = n_k\mathbb{P}_{n_k}(\phi(Y)\mathbf{1}_{\hat{\zeta}_{n_k}<Z\le\hat{\zeta}_{n_k}+h/n_k})$, $S_{k,h}^\phi = n_k\mathbb{P}_{n_k}(\phi(Y)\mathbf{1}_{\zeta_0-B_\delta/n_k<Z\le\zeta_0+(h+B_\delta)/n_k})$ and

$$T_{k,h}^\phi = n_k\mathbb{P}_{n_k}\big(\phi(Y)\mathbf{1}_{\zeta_0+B_\delta/n_k<Z\le\zeta_0+(h-B_\delta)/n_k}\big).$$

Then $\{n_k|\hat{\zeta}_{n_k} - \zeta_0| \le B_\delta\} \subset \{S_{k,h}^\phi \le \hat{T}_{k,h}^\phi \le T_{k,h}^\phi\}$, and therefore we have $\mathbf{P}(\hat{T}_{k,h}^\phi \le T_{k,h}^\phi) \wedge \mathbf{P}(S_{k,h}^\phi \le \hat{T}_{k,h}^\phi) > 1 - \delta$ for all k.

We know that $T_{k,h}^\phi \rightsquigarrow$ CPoisson$(\mu_\beta, (h - 2B_\delta)f(\zeta_0))$ and

$$\begin{aligned} S_k^\phi &\rightsquigarrow \mathrm{CPoisson}(\mu_\alpha, 2B_\delta f(\zeta_0)) \bigstar \mathrm{CPoisson}\big(\mu_\beta, (h + B_\delta)f(\zeta_0)\big) \\ &\equiv \mathrm{CPoisson}\left(\frac{B_\delta}{h + 2B_\delta}\mu_\alpha + \frac{h + B_\delta}{h + 2B_\delta}\mu_\beta, (h + 2B_\delta)f(\zeta_0)\right) \end{aligned} \tag{17}$$

as $k \to \infty$. An application of Lemma A.5 with $\mu = \nu = \mu_\beta$, $\gamma = \mu_\alpha$, $B = 4B_\delta f(\zeta_0)$, $r = \frac{1}{4}$ and $\lambda = (h - 2B_\delta)f(\zeta_0)$, shows the existence of an $h_* > 2B_\delta$ large enough so that whenever $h \ge h_*$ we can find two numbers $R_{1,h} > R_{2,h} \in \mathbb{N}$ with the property that $\underline{\lim}_{k\to\infty}\mathbf{P}(T_{k,h}^\phi < R_{2,h}) > 2\delta$ and $\underline{\lim}_{k\to\infty}\mathbf{P}(S_{k,h}^\phi \ge R_{1,h}) > 2\delta$. Thus, for $h \ge h_*$, $\mathbf{P}(T_{k,h}^\phi < R_{2,h}) > 2\delta$ and $\mathbf{P}(S_{k,h}^\phi \ge R_{1,h}) > 2\delta$ for all but a finite number of k's. Therefore, for any k large enough, $\mathbf{P}(T_{k,h}^\phi < R_{2,h}) \wedge \mathbf{P}(S_{k,h}^\phi \ge R_{1,h}) > 2\delta$. Using the fact that $\mathbf{P}(S_{k,h}^\phi \le \hat{T}_{k,h}^\phi \le T_{k,h}^\phi) > 1 - \delta$ we get that $\mathbf{P}(\hat{T}_{k,h}^\phi \le T_{k,h}^\phi < R_{2,h}) \wedge \mathbf{P}(R_{1,h} \le S_{k,h}^\phi \le \hat{T}_{k,h}^\phi) > \delta$ for all but finitely many k's. Thus, whenever $h \ge h_*$,

$$\mathbf{P}(\hat{T}_{k,h}^\phi \le T_{k,h}^\phi < R_{2,h}, \text{i.o.}) > \delta \quad \text{and} \quad \mathbf{P}(R_{1,h} \le S_{k,h}^\phi \le \hat{T}_{k,h}^\phi, \text{i.o.}) > \delta.$$

The argument relying on the Hewitt–Savage 0–1 law applied in the proof of (ii) can be used to finish this proof. A completely analogous proof applies for $\{n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\hat{\zeta}_n-h/n<Z\le\hat{\zeta}_n})\}_{n=1}^\infty$. $\square$

Acknowledgments. We would like to thank Souvik Ghosh for his insightful remarks on the proof of Lemma 4.3. We also appreciate the helpful comments made by Moulinath Banerjee two anonymous referees.

SUPPLEMENTARY MATERIAL

Supplement to "Change-point in stochastic design regression and the bootstrap" (DOI: 10.1214/11-AOS874SUPP; .pdf). The supplementary file contains a longer version of this paper with all the technical details which were excluded in the present version due to their length.

REFERENCES

[1] ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245

[2] BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31. MR1441142

[3] BILLINGSLEY, P. (1986). *Probability and Measure*. Wiley, New York. MR0830424

[4] BOSE, A. and CHATTERJEE, S. (2001). Generalised bootstrap in non-regular M-estimation problems. *Statist. Probab. Lett.* **55** 319–328. MR1867535

[5] DEMPFLE, A. and STUTE, W. (2002). Nonparametric estimation of a discontinuity in regression. *Statist. Neerlandica* **56** 233–242. MR1916322

[6] DUDLEY, R. M. (2002). *Real Analysis and Probability*. Cambridge Univ. Press, Cambridge. MR1932358

[7] DÜMBGEN, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.* **19** 1471–1495. MR1126333

[8] EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics* **38**. SIAM, Philadelphia, PA. MR0659849

[9] ETHIER, S. and KURTZ, T. (2005). *Markov Processes, Characterization and Convergence*. Wiley, New York. MR0838085

[10] FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228. MR0630104

[11] GIJBELS, I., HALL, P. and KNEIP, A. (1999). On the estimation of jump point in smooth curves. *Ann. Inst. Statist. Math.* **51** 231–251. MR1707773

[12] GIJBELS, I., HALL, P. and KNEIP, A. (2004). Interval and band estimation for curves with jumps. *J. Appl. Probab.* **41** 65–79. MR2057566

[13] HŬSKOVÁ, M. and KIRCH, C. (2008). Bootstrapping confidence intervals for the change-point of time series. *J. Time Ser. Anal.* **29** 947–972. MR2464948

[14] 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. Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212

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

[16] KOSOROK, M. R. and SONG, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.* **35** 957–989. MR2341694

[17] KOUL, H., QIAN, L. and SURGAILIS, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Process. Appl.* **103** 123–154. MR1947962

[18] LAN, Y., BANERJEE, M. and MICHAILIDIS, G. (2009). Change-point estimation under adaptive sampling. *Ann. Statist.* **37** 1752–1791. MR2533471

[19] LEE, S. M. S. and PUN, M. C. (2006). On m out of n bootstrapping for nonstandard M-estimation with nuisance parameters. *J. Amer. Statist. Assoc.* **101** 1185–1197.

[20] MÖHLE, M. (2005). Convergence results for compound Poisson distributions and applications to the standard Luria–Delbrück distribution. *J. Appl. Probab.* **42** 620–631. MR2157509

[21] NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295. MR0293706

[22] POLITIS, D., ROMANO, J. and WOLF, W. (1999). *Subsampling*. Springer, New York. MR1707286

[23] PONS, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold covariate. *Ann. Statist.* **31** 442–463. MR1983537

[24] SCOTT, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York. MR1191168

[25] SEIJO, E. and SEN, B. (2011). Supplement to "Change-point in stochastic design regression and the bootstrap." DOI:10.1214/11-AOS874SUPP.

[26] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880

[27] SHAO, J. and TU, D. (1995). *The Jackknife and Bootstrap*. Springer, New York. MR1351010

[28] SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184. MR0471166

[29] WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1295. MR0868303

DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
1255 AMSTERDAM AVENUE
NEW YORK, NEW YORK 10027
USA
E-MAIL: emilio@stat.columbia.edu
bodhi@stat.columbia.edu
URL: http://www.stat.columbia.edu/~emilio
http://www.stat.columbia.edu/~bodhi