INFERENCE UNDER RANDOM LIMIT BOOTSTRAP MEASURES

Giuseppe Cavaliere
Department of Economics, University of Bologna
Department of Economics, University of Exeter

Iliyan Georgiev
Department of Economics, University of Bologna
Nova School of Business and Economics, Universidade Nova de Lisboa

First draft: July 2018; Revised: December 2019; This version: June 2, 2020

Asymptotic bootstrap validity is usually understood as consistency of the distribution of a bootstrap statistic, conditional on the data, for the unconditional limit distribution of a statistic of interest. From this perspective, randomness of the limit bootstrap measure is regarded as a failure of the bootstrap. We show that such limiting randomness does not necessarily invalidate bootstrap inference if validity is understood as control over the frequency of correct inferences in large samples. We first establish sufficient conditions for asymptotic bootstrap validity in cases where the unconditional limit distribution of a statistic can be obtained by averaging a (random) limiting bootstrap distribution. Further, we provide results ensuring the asymptotic validity of the bootstrap as a tool for conditional inference, the leading case being that where a bootstrap distribution estimates consistently a conditional (and thus, random) limit distribution of a statistic. We apply our framework to several inference problems in econometrics, including linear models with possibly non-stationary regressors, CUSUM statistics, conditional Kolmogorov-Smirnov specification tests and tests for constancy of parameters in dynamic econometric models.

Keywords: Bootstrap, random measures, weak convergence in distribution, asymptotic inference.

Giuseppe Cavaliere: giuseppe.cavaliere@unibo.it
Iliyan Georgiev: i.georgiev@unibo.it

We are grateful to the Co-Editor and three anonymous referees for the many constructive comments and suggestions on earlier versions of the paper. We have also benefited from discussions and feedback from Matias Cattaneo, Graham Elliott, Paola Giuliano, Michael Jansson, Søren Johansen, Ye Lu, Adam McCloskey, Marcelo Moreira, Rasmus Søndergaard Pedersen, Anders Rahbek, Mervyn Silvapulle, Michael Wolf, seminar participants at Boston U, Cambridge, ETH, Exeter, FGV São Paulo, Genève, Hitotsubashi, Kyoto, LSE, Melbourne, Oxford, PUC Rio, Queen's, QUT, SMU, Tilburg, UCL, UPF, Verona, Vienna, as well as participants to the 7th Italian Congress of Econometrics and Empirical Economics (ICEEE), the 28th $(EC)^2$ Conference, the 14th SETA meeting, the NBER/NSF 2018 Time Series Conference and the IX_t Zaragoza Workshop in Time Series Econometrics. We also thank Edoardo Zanelli and Lorenzo Zannin for research assistance. Financial support from the University of Bologna, ALMA IDEA 2017 grants, and from the Danish Council for Independent Research (DSF Grant 7015-00028), is gratefully acknowledged.

1 Introduction

Consider a data sample D_n of size n and a statistic $\tau_n := \tau_n(D_n)$, say a test statistic or a parameter estimator, possibly normalized. Interest is in a distributional approximation of τ_n. Let a bootstrap procedure generate a bootstrap analogue τ_n^* of τ_n; i.e., computed on a bootstrap sample. Assume that τ_n converges in distribution to a non-degenerate random variable [rv], say τ. In classic bootstrap inference, asymptotic bootstrap validity is understood and established as convergence in probability (or almost surely) of the cumulative distribution function [cdf] of the bootstrap statistic τ_n^* conditional on the data D_n, say F_n^*, to the unconditional cdf of τ, say F. This convergence, along with continuity of F, implies by Polya's theorem that $\sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \to 0$, in probability (or almost surely).

In many applications, however, the bootstrap statistic τ_n^* may possess, conditionally on the data, a *random* limit distribution. Cases of random bootstrap limit distributions appear in various areas of econometrics and statistics; for instance, they are documented for infinite-variance processes (Athreya, 1987; Aue, Berkes and Horváth, 2008), time series with unit roots (Basawa et al., 1991; Cavaliere, Nielsen and Rahbek, 2015), parameters on the boundary of the parameter space (Andrews, 2000), block resampling methods under fixed-b asymptotics (Lahiri, 2001; Shao and Politis, 2013), cube-root consistent estimators (Sen, Banerjee and Woodroofe, 2010; Cattaneo, Jansson and Nagasawa, 2020), Hodges-LeCam superefficient estimators (Beran, 1997). In most of these cases, the occurrence of a random limit distribution for the bootstrap statistic τ_n^* given the data – in contrast to a non-random limit of the unconditional distribution of the original statistic τ_n – is taken as evidence of bootstrap failure[1].

In this paper we show that randomness in the limit distribution of a bootstrap statistic need not invalidate bootstrap inference. On the contrary, although the bootstrap no longer estimates the limiting unconditional distribution of the statistic of interest, it may still deliver hypothesis tests (or confidence intervals) with the desired null rejection probability (or coverage probability) when the sample size diverges. This happens because asymptotic control over the frequency of wrong inferences can be guaranteed by the asymptotic distributional uniformity of the bootstrap p-values, which in its turn can occur without the convergence in probability (or almost surely) of the bootstrap cdf F_n^* of τ_n^* to the asymptotic cdf F of τ.

Therefore, in cases where the limit bootstrap distribution is random, our analysis focuses on the asymptotics of the bootstrap p-value $p_n^* := F_n^*(\tau_n)$. We define '(unconditional) bootstrap validity' as the fact that

$$P(p_n^* \leq q) \to q \tag{1.1}$$

for all $q \in (0, 1)$. Interest in this property is not new in the literature (see, e.g., Hansen, 1996, and Lockhart, 2012, among others). When the limit bootstrap distribution is

[1] For instance, Athreya (1987) writes: 'If the bootstrap were to be successful here, then F_n^* [our notation] should converge to F [our notation] in distribution. However, this is not the case. There is a random limit.' (p.725) and 'constructing confidence intervals on the basis of a Monte Carlo simulation of the bootstrap could lead to misleading results' (p.728). Similar considerations appear in, *inter alia*, Basawa et al. (1991), Lahiri (2001), Aue et al. (2008).

random, however, the existing research provides mostly counterexamples to property (1.1) (e.g., Shao and Politis, 2013), whereas general sufficient conditions for bootstrap validity in the sense of (1.1) have not been studied.

Our first set of results provides such sufficient conditions. Classic results for bootstrap validity when the limit bootstrap measure is not random can be obtained as special cases. The main requirement in our results is that the unconditional limit distribution of τ_n should be an average of the random limit distribution of τ_n^* given the data.

It is often the case that bootstrap validity can be addressed through the lens of a conditioning argument. In this regard, our second set of results concerns the possibility that, for a sequence of random elements X_n, it holds that the bootstrap p-value is uniformly distributed in large sample *conditionally* on X_n:

$$P(p_n^* \leq q|X_n) \xrightarrow{p} q \tag{1.2}$$

for all $q \in (0, 1)$. This property, that we call 'bootstrap validity conditional on X_n', implies unconditional validity in the sense of (1.1). Moreover, conditional bootstrap validity given X_n implies that the bootstrap replicates asymptotically the property of conditional tests and confidence intervals to have, conditionally on X_n, constant null rejection probability and coverage probability, respectively (for further roles of conditioning in inference, like the relevance of the drawn inferences and information recovery, see Reid, 1995, and the references therein). The leading case where we show (1.2) to hold – under regularity conditions that will be discussed in the paper – is that where the (random) limit of the conditional distribution of τ_n given X_n matches the (random) limit distribution of the bootstrap statistic. The idea of comparing the limit bootstrap distribution with the limit of a conditional distribution of a statistic of interest was put forward by Lepage and Podgórski (1996), but was not recast in terms of bootstrap p-values as in eq. (3.2). The use of a conditioning argument to establish (1.1) and (1.2) can be found in Cavaliere, Georgiev and Taylor (2013) and Georgiev et al. (2019), respectively. Their results follow as special cases of those provided in this paper.

When dealing with random limit distributions, the usual convergence concept employed to establish bootstrap validity, i.e. weak convergence in probability, can only be used in some very special cases. Instead, our formal discussion makes extensive use of the probabilistic concept of weak convergence of random measures; see e.g. Kallenberg (2017, Ch.4). To our knowledge, in the bootstrap context this concept has so far been mostly used to obtain negative results of lack of validity for specific bootstrap procedures (see above), rather than positive validity results, as we do here. As an ingredient of our analysis, we also present some novel results on the weak convergence of conditional expectations.

To provide motivation for the practical relevance of our results, we initially illustrate them by using a simple linear model with either stationary or non-stationary regressors, and later we analyze three well-known cases in the econometric literature where the bootstrap features a random limit distribution. The first is a standard CUSUM-type test of the i.i.d. property for a random sequence with infinite variance. This is a case where the limit distribution of the CUSUM statistic depends on unknown nuisance parameters (e.g., the tail index) and bootstrap or permutation tests fail to estimate this distribution consistently. We argue that a simple bootstrap based on permutations, albeit having a random limit distribution and hence being invalid in the usual sense, provides *exact* conditional inference and hence is also unconditionally valid in the sense of (1.1).

The second application considers a Kolmogorov-Smirnov-type test for correct specification of the conditional distribution of a response variable given a vector of covariates. Andrews (1997) considers a parametric bootstrap implementation where the covariates are kept fixed across bootstrap samples. While in the independent case the limit of the bootstrap distribution is non-random, this is not the case in general. Using our theory we discuss conditions for validity of the bootstrap within this framework.

Finally, we consider the well-known and widely applied case of bootstrap implementations of '$\sup F$' tests for parameter constancy in regressions models where the regressors could be non-stationary, with the latter including the case of regressors subject to (possibly random) structural change. As in Hansen (2000), see also Hall (1992, p.170), in the resampling process forming the bootstrap sample, it appears natural to take the design matrix as fixed across the bootstrap repetitions. Under a set of assumptions proposed by Hansen (2000), we argue that the fixed-regressor bootstrap '$\sup F$' statistic has a random limit distribution, thus invalidating previous claims in the literature that the bootstrap is consistent for the unconditional limit distribution of the original '$\sup F$' test statistic. We then provide conditions under which the fixed-regressor bootstrap is valid, unconditionally and conditionally on the chosen set of regressors.

Structure of the paper

The paper is organized as follows. In Section 2 we outline the central concepts and ideas using a simple linear regression model. The main theoretical results are presented in Section 3. Section 4 contains the three applications of the theory, whereas Section 5 concludes. The paper has two Appendices. In Appendix A we collect some results on weak convergence in distribution which are useful to prove the main theorems and develop the applications. Appendix B contains the proofs of the main theorems. Additional material and proofs are given in the accompanying supplement (Cavaliere and Georgiev, 2020). Sections, equations, etc., numbered S.x can be found there.

Notation and definitions

We use the following notation throughout. The spaces of càdlàg functions $[0,1] \to \mathbb{R}^n$, $[0,1] \to \mathbb{R}^{m\times n}$ and $\mathbb{R} \to \mathbb{R}$ (all equipped with the respective Skorokhod J_1-topologies; see Kallenberg, 1997, Appendix A2), are denoted by $\mathscr{D}_n$, $\mathscr{D}_{m\times n}$ and $\mathscr{D}(\mathbb{R})$, respectively; for the first one, when $n=1$ the subscript is suppressed. Integrals are over $[0,1]$ unless otherwise stated, Φ is the standard Gaussian cdf, $U(0,1)$ is the uniform distribution on $[0,1]$ and $\mathbb{I}_{\{\cdot\}}$ is the indicator function. If F is a (random) cdf, F^{-1} stands for the right-continuous generalized inverse, i.e., $F^{-1}(u) := \sup\{v \in \mathbb{R} : F(v) \leq u\}$, $u \in \mathbb{R}$. Unless differently specified, limits are for $n \to \infty$.

Polish (i.e., complete and separable metric) spaces are always equipped with their Borel σ-algebras. Throughout, we assume that all the considered random elements are Polish-space valued. For random elements of a Polish space, the existence of regular conditional distributions is guaranteed and we assume without loss of generality that conditional probabilities are regular (Kallenberg, 1997, Theorem 5.3). Equality of conditional distributions is understood in the almost sure [a.s.] sense and, for random cdf's as random elements of $\mathscr{D}(\mathbb{R})$, equalities are up to indistinguishability.

Let $\mathcal{C}_b(\mathcal{S})$ be the set of all continuous and bounded real-valued functions on a metric space $\mathcal{S}$. For random elements Z, Z_n ($n \in \mathbb{N}$) of a metric space $\mathcal{S}_Z$, we employ the usual notation $Z_n \overset{w}{\to} Z$ for the property that the distribution of Z_n weakly converges to the distribution of Z, defined by the convergence $E\{g(Z_n)\} \to E\{g(Z)\}$ for all $g \in \mathcal{C}_b(\mathcal{S}_Z)$. For random elements (Z, X), (Z_n, X_n) of the metric spaces $\mathcal{S}_Z \times \mathcal{S}$ and $\mathcal{S}_Z \times \mathcal{S}_n$ ($n \in \mathbb{N}$), and defined on a common probability space, we denote by $Z_n|X_n \overset{w}{\to}_p Z|X$ (resp. $Z_n|X_n \overset{w}{\to}_{a.s.} Z|X$) the fact that $E\{g(Z_n)|X_n\} \to E\{g(Z)|X\}$ in probability (resp. a.s.) for all $g \in \mathcal{C}_b(\mathcal{S}_Z)$. In the special case where $E\{g(Z_n)|X_n\} \overset{w}{\to} E\{g(Z)\}$ in probability (resp. a.s.) for all $g \in \mathcal{C}_b(\mathcal{S}_Z)$, we write $Z_n|X_n \overset{w}{\to}_p Z$ (resp. $Z_n|X_n \overset{w}{\to}_{a.s.} Z$). In such a case the weak limit (in probability or a.s.) of the random conditional distribution $Z_n|X_n$ is the non-random distribution of Z, thus reducing our definition to the one of weak convergence in probability (resp. a.s.) usually employed in the bootstrap literature.

In order to deal with random limit measures, we need a further convergence concept. For $(Z, X), (Z_n, X_n)$ ($n \in \mathbb{N}$) defined on possibly different probability spaces, we denote by $Z_n|X_n \overset{w}{\to}_w Z|X$ the fact that $E\{g(Z_n)|X_n\} \overset{w}{\to} E\{g(Z)|X\}$ for all $g \in \mathcal{C}_b(\mathcal{S}_Z)$ and label it 'weak convergence in distribution'. It coincides with the probabilistic concept of weak convergence of random measures (here, of the random conditional distributions $Z_n|X_n$; see Kallenberg, 2017, Ch.4). Whenever Z_n and Z are rv's and the conditional distribution of Z given X is *diffuse* (non-atomic), this is equivalent to the weak convergence $P(Z_n \leq \cdot|X_n) \overset{w}{\to} P(Z \leq \cdot|X)$ of the random cdf's as random elements of $\mathscr{D}(\mathbb{R})$ (see Kallenberg, 2017, Theorem 4.20). Finally, on probability spaces where both the data D_n and the auxiliary variates used in the construction of the bootstrap data are defined, we use $Z_n \overset{w^*}{\to}_p Z|X$ (resp. $\overset{w^*}{\to}_{a.s}$, $\overset{w^*}{\to}_w$) interchangeably with $Z_n|D_n \overset{w}{\to}_p Z|X$ (resp. $\overset{w}{\to}_{a.s}$, $\overset{w}{\to}_w$), and write $P^*(\cdot)$ for $P(\cdot|D_n)$.

2 A linear regression example

In this section we provide an overview of the main results established in the sections to follow, and the concepts employed, by using a simple linear regression model. Further applications will be given in Section 4. We observe that even for this basic model bootstrap statistics may have a random limit distribution. Then, we show that convergence of the bootstrap statistic to a random limit may imply bootstrap validity in the unconditional sense of eq. (1.1). Finally, we illustrate the possibility that bootstrap inference may have a conditional interpretation.

2.1 Model, bootstrap and random limit bootstrap measures

Assume that the data are given by $D_n := \{y_t, x_t\}_{t=1}^n$ and consider the linear model

$$y_t = \beta x_t + \varepsilon_t \qquad (t = 1, 2, ..., n) \tag{2.1}$$

where x_t, y_t are scalar rv's and ε_t are unobservable zero-mean errors with $\omega_\varepsilon := \mathrm{Var}(\varepsilon_t) \in (0, \infty)$, $t = 1, ..., n$. Assume that $M_n := \sum_{t=1}^n x_t^2 > 0$ a.s. for all n; further assumptions will be introduced gradually. Interest is in inference on β based on $T_n := \hat{\beta} - \beta$, with $\hat{\beta}$ the OLS estimator of β; for instance, a confidence interval or a test of a null hypothesis of the form $\mathsf{H}_0 : \beta = 0$.

The classic (parametric) fixed-design bootstrap, see e.g. Hall (1992), entails generating a bootstrap sample $\{y_t^*, x_t\}_{t=1}^n$ as

$$y_t^* = \hat{\beta} x_t + \hat{\omega}_\varepsilon^{1/2} \varepsilon_t^* \qquad (t = 1, 2, ..., n) \tag{2.2}$$

where $\{\varepsilon_t^*\}_{t=1}^n$ are i.i.d. $N(0, 1)$, independent of the original data, and $\hat{\omega}_\varepsilon$ is an estimator of ω_ε, e.g., the residual variance $n^{-1} \sum_{t=1}^n (y_t - \hat{\beta} x_t)^2$. The OLS estimator of β from the bootstrap sample is denoted by $\hat{\beta}^*$ and, conditionally on the original data, $T_n^* := \hat{\beta}^* - \hat{\beta} \sim N\left(0, \hat{\omega}_\varepsilon M_n^{-1}\right)$. As is standard, the distribution of T_n is approximated by the distribution of T_n^* conditional on the data. With F_n^* denoting the cdf of T_n^* under P^*, the bootstrap p-value is given by $p_n^* := F_n^*(T_n)$.

Remark 2.1 A special case where the ensuing bootstrap inference is exact in finite samples, such that p_n^* is uniformly distributed for finite n, obtains when the original ε_t's are $N(0, \omega_\varepsilon)$, independent of $X_n := \{x_t\}_{t=1}^n$, and ω_ε is known to the econometrician (hence $\hat{\omega}_\varepsilon = \omega_\varepsilon$). Then the conditional distribution of T_n^* given the data D_n and the distribution of the original statistic T_n *conditional* on the regressor X_n (equivalently, on the ancillary statistic M_n), are a.s. equal to each other and to the conditional distribution $N\left(0, \omega_\varepsilon M_n^{-1}\right)|M_n$. Put differently,

$$F_n^*(u) := P(T_n^* \le u | D_n) = P(T_n \le u | X_n) = \Phi(\omega_\varepsilon^{-1/2} M_n^{1/2} u),\ u \in \mathbb{R}.$$

Then, as $\omega_\varepsilon^{-1/2} M_n^{1/2} T_n | M_n \sim N(0, 1)$, it is straightforward that in this special case bootstrap inference is exact: $p_n^* = F_n^*(T_n) = \Phi(\omega_\varepsilon^{-1/2} M_n^{1/2} T_n) \overset{d}{=} \Phi(N(0, 1)) \sim U(0, 1)$, and that this result also holds conditionally on M_n: $p_n^* | M_n \sim U(0, 1)$. □

Although bootstrap inference is not exact in general, it may still be asymptotically valid. To show this, we distinguish between the cases of a stationary and a non-stationary regressor x_t. It is the second case that anticipates the main results of the paper. We assume $\hat{\omega}_\varepsilon \xrightarrow{p} \omega_\varepsilon$ throughout.

2.1.1 Classic bootstrap validity when the regressor is stationary

Suppose initially that $\{x_t\}_{t \in \mathbb{N}}$ is weakly stationary and $n^{-1} M_n \xrightarrow{p} M := E x_1^2 > 0$. Define $\tau_n := n^{1/2}(\hat{\beta} - \beta)$ and $\tau_n^* := n^{1/2}(\hat{\beta}^* - \hat{\beta})$; the bootstrap p-values based on (τ_n, τ_n^*) and (T_n, T_n^*) are identical. The distribution of the bootstrap statistic τ_n^* conditional on the original data D_n satisfies

$$P^*(\tau_n^* \le u) = \Phi(n^{-1/2} \hat{\omega}_\varepsilon^{-1/2} M_n^{1/2} u) \xrightarrow{p} \Phi(\omega_\varepsilon^{-1/2} M^{1/2} u),\ u \in \mathbb{R}. \tag{2.3}$$

Hence, $\tau_n^* \xrightarrow{w^*}_p \tau \sim N(0, \omega_\varepsilon M^{-1})$ and the limit distribution is non-random.

If the initial assumptions are strengthened such that a central limit theorem [CLT] holds for $\{x_t \varepsilon_t\}_{t \in \mathbb{N}}$; that is, $n^{-1/2} \sum_{t=1}^n x_t \varepsilon_t \xrightarrow{w} N(0, \omega_\varepsilon M)$, then it further holds that $\tau_n \xrightarrow{w} \tau \sim N(0, \omega_\varepsilon M^{-1})$. Hence, the bootstrap distribution of τ_n^* consistently estimates the unconditional limit distribution of τ_n in the usual sense that $\sup_{u \in \mathbb{R}} |P^*(\tau_n^* \le u) - P(\tau \le u)| \xrightarrow{p} 0$, by Polya's theorem. As the limit cdf is continuous, the p-value p_n^* associated with (τ_n, τ_n^*) is asymptotically uniformly distributed and (1.1) holds.

2.1.2 Random limit bootstrap measures when the regressor is non-stationary

Suppose now that $\{x_t\}_{t\in\mathbb{N}}$ is such that, for some constant α, $n^{-\alpha}M_n \xrightarrow{w} M$, with $M > 0$ a.s. having a non-degenerate distribution. A well-known special case is that where x_t is a finite-variance random walk and $\alpha = 2$. Redefine $\tau_n := n^{\alpha/2}(\hat{\beta} - \beta)$ and $\tau_n^* := n^{\alpha/2}(\hat{\beta}^* - \hat{\beta})$; bootstrap p-values remain unchanged. Now the bootstrap distribution of τ_n^*, conditional on the data, remains random in the limit. Specifically, by the continuous mapping theorem [CMT],

$$P^*(\tau_n^* \leq u) = \Phi(n^{-\alpha/2}\hat{\omega}_\varepsilon^{-1/2}M_n^{1/2}u) \xrightarrow{w} \Phi(\omega_\varepsilon^{-1/2}M^{1/2}u),\ u \in \mathbb{R}, \tag{2.4}$$

which is a random cdf. In terms of weak convergence in distribution, this amounts to

$$\tau_n^* \xrightarrow{w^*}_w N(0, \omega_\varepsilon M^{-1})\big|\, M. \tag{2.5}$$

As a result, with τ_n^* and M generally defined on different probability spaces, weak convergence in probability of τ_n^* does not occur. Moreover, whatever the (unconditional) limit distribution of τ_n is, provided that it exists, $P(\tau_n \leq u)$, $u \in \mathbb{R}$, will tend to a deterministic cdf. Therefore, the bootstrap cannot estimate consistently the limit distribution of τ_n and it cannot hold that $\sup_{u\in\mathbb{R}} |P^*(\tau_n^* \leq u) - P(\tau \leq u)| \xrightarrow{p} 0$. Nevertheless, bootstrap inference need not become meaningless, as it may even be exact (see Remark 2.1). We proceed, therefore, to identify in what sense bootstrap inference could remain meaningful.

2.2 Bootstrap validity

Within the framework of the linear regression model, we discuss two concepts of bootstrap validity in the case of a random limit bootstrap measure. These are employed to interpret the bootstrap as a tool for unconditional or conditional inference.

2.2.1 Unconditional bootstrap validity

Under the assumption in Section 2.1.2, consider the random-walk special case, where $x_t := \sum_{s=1}^t \eta_s$ with $e_t := (\varepsilon_t, \eta_t)'$ forming a stationary, ergodic and conditionally homoskedastic martingale difference sequence [mds] with p.d. variance matrix $\Omega := \text{diag}\{\omega_\varepsilon, \omega_\eta\}$.[2] Then, for $\beta \neq 0$ eq. (2.1) is an instance of a cointegration regression. It holds that $(n^{-1/2}\sum_{t=1}^{\lfloor n\cdot\rfloor} e_t', n^{-1}\sum_{t=1}^n x_{t-1}\varepsilon_t) \xrightarrow{w} (B_\varepsilon, B_\eta, \int B_\eta dB_\varepsilon)$ in $\mathscr{D}_2 \times \mathbb{R}$, where $(B_\varepsilon, B_\eta)'$ is a bivariate Brownian motion with covariance matrix Ω; see Theorem 2.4 of Chan and Wei (1988). Moreover, $n^{-2}M_n \xrightarrow{w} M := \int B_\eta^2$ by the CMT, jointly with the convergence to a stochastic integral above, so that the assumption in Section 2.1.2 holds with $\alpha = 2$ and

$$\tau_n := n(\hat{\beta} - \beta) \xrightarrow{w} \left(\int B_\eta^2\right)^{-1} \int B_\eta dB_\varepsilon \sim N(0, \omega_\varepsilon M^{-1}), \tag{2.6}$$

[2] Non-diagonal Ω could be handled by augmenting the estimated regression with Δx_t, leading to no qualitative differences from the case of diagonal Ω.

the limit being (by independence of B_η and B_ε) a variance mixture of normals, with mixing variable M^{-1} and cdf $\int_{\mathbb{R}} \Phi(\omega_\varepsilon^{-1/2} M^{1/2} u) dP(M)$.

A comparison between the limit distributions of τ_n^* and τ_n, resp. in (2.5) and (2.6), shows that the bootstrap mimics a component of the mixture limit distribution of τ_n, since the limit distribution of τ_n can be recovered by integrating over M the conditional limit distribution of τ_n^* given the data. This turns out to be sufficient for unconditional bootstrap validity in the sense of eq. (1.1). A direct argument is as follows: the bootstrap p-value $p_n^* := P^*(\tau_n^* \le \tau_n)$ satisfies, by the CMT,

$$p_n^* = \Phi(\hat{\omega}_\varepsilon^{-1/2} M_n^{1/2} (\hat{\beta} - \beta)) \xrightarrow{w} \Phi((\omega_\varepsilon \textstyle\int B_\eta^2)^{-1/2} \int B_\eta dB_\varepsilon) \stackrel{d}{=} \Phi(N(0,1)) \sim U(0,1). \tag{2.7}$$

Thus, when inference on β is based on the distribution of τ_n^* conditional on the data, the large-sample frequency of wrong inferences can be controlled.

2.2.2 Conditional bootstrap validity

In the case of unconditional bootstrap validity, it may be possible to find an interpretation of bootstrap inference as also valid in the sense of (1.2), *i.e.* conditionally on some X_n defined on the probability space of the original data D_n (for instance, but not necessarily, the regressor $X_n := \{x_t\}_{t=1}^n$).

In the linear regression case considered here, conditional bootstrap validity with respect to the regressor X_n can be obtained under a tightening of our previous assumptions such that the invariance principle $n^{-1/2} \sum_{t=1}^{\lfloor n\cdot \rfloor} e_t \xrightarrow{w} (B_\varepsilon, B_\eta)'$ holds *conditionally* (on X_n for finite n and on B_η in the limit, in the sense of weak convergence in distribution). A sufficient condition for the conditional invariance principle is that, additionally to the assumptions on e_t in Section 2.2.1, ε_t is an mds with respect to $\mathcal{G}_t = \sigma(\{\varepsilon_s\}_{s=-\infty}^t \cup \{\eta_s\}_{s\in\mathbb{Z}})$, and that $n^{-1} \sum_{t=1}^n E(\varepsilon_t^2 | \{\eta_s\}_{s\in\mathbb{Z}}) \to \omega_\varepsilon$ a.s. (see the proof of Theorem 2 in Rubshtein, 1996). Then, by using Theorem 3 of Georgiev et al. (2019), it follows that

$$\tau_n | X_n \xrightarrow{w}_w N(0, \omega_\varepsilon M^{-1}) \big| M,$$

which compared to (2.5) shows that the distribution of τ_n^* conditional on the data estimates consistently the random limit distribution of τ_n conditional on the regressor X_n. This fact is stated more precisely in Remark 3.9 where it is concluded that $p_n^* | X_n \xrightarrow{w}_p U(0,1)$, *i.e.*, the bootstrap is valid conditionally on the regressor.

2.2.3 A numerical illustration

The result in Section 2.2.2 implies that unconditional bootstrap validity can sometimes be established by means of a conditioning argument; for example, by showing validity conditional on the regressor X_n. To illustrate, in Figure 1, panels (i) and (ii), we summarize for two different data generating processes [DGPs] the cdf's of $p_n^* | X_n$ across $M = 1,000$ independent realizations of X_n for samples of size $n = 10$ (upper panels) and $n = 1,000$ (lower panels). Specifically, the DGP used for panel (i) is based on i.i.d. shocks, while the one for (ii) features ARCH-type shocks (details are reported in Section S.5).

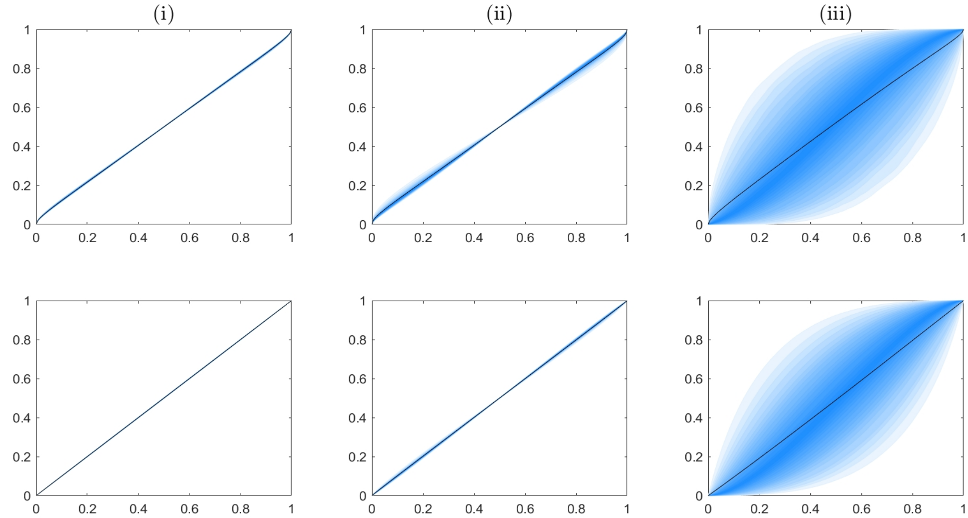

FIGURE 1: *Fan chart of the simulated cdfs (conditional on X_n) of the bootstrap p-values for the three DGPs (i)–(iii) and* $n = 10$ *(upper panels),* 1000 *(lower panels).*

In both cases, the conditions of Section 2.2.2 are satisfied. For both DGPs, the conditional cdf's of p_n^* given X_n are, as expected, close to the 45° line, which corresponds to the implied asymptotic $U(0,1)$ distribution. Unconditional validity follows accordingly. Nevertheless, unconditional validity may also hold without validity conditional on an apparently 'natural' conditioning variable X_n, like the regressor in a fixed-regressor bootstrap design. For instance, suppose that for the DGP in Sections 2.2.1 and 2.2.2 it holds that $\eta_t = \xi_t(1 + \mathbb{I}_{\{\varepsilon_t<0\}})$, with $\{\varepsilon_t\}$ and $\{\xi_t\}$ two independent i.i.d. sequences of zero-mean, unit-variance rv's. Since η_t is informative about the sign of ε_t, the ε_t's conditionally on their own past and the regressor X_n do not form an mds. It is shown in Section S.3, eq. (S.9), that this endogeneity fact, not replicated in the bootstrap world, induces the original statistic τ_n to satisfy

$$\tau_n|X_n \xrightarrow{w}_w M^{-1/2}(\omega_{\varepsilon|\eta}^{1/2}\xi_1 + (1-\omega_{\varepsilon|\eta})^{1/2}\xi_2)\,|(M,\xi_2)\,, \tag{2.8}$$

where $\omega_{\varepsilon|\eta} := E\{\text{Var}(\varepsilon_s|\eta_s)\} \in (0,1)$, and M, ξ_1, ξ_2 are jointly independent with $\xi_i \sim N(0,1), i = 1,2$. The limit in (2.8) contains more randomness (through ξ_2) than the bootstrap limit in eq. (2.5), thus resulting in a random limit for the distribution of the bootstrap p-value p_n^* conditional on X_n; see panel (iii) of Figure 1, where for this DGP the cdf's of $p_n^*|X_n$ are reported for 1,000 realizations of X_n. These cdf's display substantial dispersion around the 45° line, and this feature does not vanish as n increases. However, and in agreement with the earlier discussion, their unconditional average (plotted in black) is very close to the 45° line, showing indeed unconditional validity of the bootstrap. This follows because $e_t := (\varepsilon_t, \eta_t)'$ is a zero-mean i.i.d. sequence with a diagonal covariance matrix and $p_n^* \xrightarrow{w} U(0,1)$ as derived in Section 2.2.1.

REMARK 2.2 Although not valid conditionally on the regressor X_n, in the previous example the bootstrap may be valid conditionally on a non-trivial function of the regressor. See, in particular, Section 3.3 and Remark 3.10 therein. □

3 Main results

We provide general conditions for bootstrap validity in cases where a bootstrap statistic conditionally on the data possesses a random limit distribution. Before all else, we formally distinguish between two concepts of bootstrap validity.

3.1 Definitions

The following definition employs the bootstrap p-value as a summary indicator of the accuracy of bootstrap inferences (see also Remarks 3.2 and 3.3 below). The original and the bootstrap statistic are denoted by τ_n and τ_n^*, respectively.

Definition 1 *Let $\tau_n := \tau_n(D_n)$ and $\tau_n^* := \tau_n^*(D_n, W_n^*)$, $n \in \mathbb{N}$, where D_n denotes the data whereas W_n^* are auxiliary variates defined jointly with D_n on a possibly extended probability space. Let $p_n^* := P(\tau_n^* \leq \tau_n | D_n)$ be the bootstrap p-value.*

We say that the bootstrap based on τ_n and τ_n^* is valid unconditionally if p_n^* is asymptotically $U(0,1)$ distributed:

$$P(p_n^* \leq q) \to q \text{ for all } q \in (0,1), \tag{3.1}$$

where $P(\cdot)$ denotes probability w.r.t. the distribution of D_n.

Let further X_n be a random element defined on the probability space of D_n and W_n^*. We say that the bootstrap based on τ_n and τ_n^* is valid conditionally on X_n if p_n^* is asymptotically $U(0,1)$ distributed conditionally on X_n :

$$P(p_n^* \leq q | X_n) \xrightarrow{p} q \text{ for all } q \in (0,1), \tag{3.2}$$

where $P(\cdot|X_n)$ is determined up to a.s. equivalence by the distribution of (D_n, X_n).

Remark 3.1 Bootstrap validity conditionally on some X_n implies unconditional validity, by the dominated convergence theorem. In applications, therefore, the discussion of conditional validity may represent an intermediate step to assess unconditional validity.

Remark 3.2 The validity properties in Definition 1 ensure correct asymptotic null rejection probability, unconditionally or conditionally on some X_n, for bootstrap hypothesis tests which reject the null when the bootstrap p-value p_n^* does not exceed a chosen nominal level, say $\alpha \in (0,1)$. If $P(\tau_n^* \leq \cdot | D_n)$ converges weakly in $\mathscr{D}(\mathbb{R})$ to a sample-path continuous random cdf, then correct asymptotic null rejection probability is ensured also for bootstrap tests rejecting the null hypothesis when $\tilde{p}_n^* := P(\tau_n^* \geq \tau_n | D_n) \leq \alpha$ (for an application, see Section 4.3).

Remark 3.3 Validity as in Definition 1 has also implications on the properties of bootstrap (percentile) confidence sets. Suppose, for instance, that T_n is an estimator of a population (scalar) parameter, whose true value is denoted by θ_0, and assume for simplicity that τ_n is of the form $\tau_n = \rho(n)(T_n - \theta_0)$, where $\rho(n)$ is a normalizing factor such that τ_n has a non-degenerate limiting distribution (see Horowitz, 2001, p.3174). Its bootstrap analog is denoted by τ_n^*, and we assume that the bootstrap is valid in the sense of (3.1). Interest is in constructing a right-sided confidence interval for θ_0, with (asymptotic) coverage $1-\alpha \in (0,1)$, using a simple bootstrap percentile method. With

$F_n^*(x) := P(\tau_n^* \leq x|D_n)$, let $q_n^*(1-\alpha) := \inf\{x \in \mathbb{R} : F_n^*(x) \geq 1-\alpha\}$ be the $(1-\alpha)$ quantile of the bootstrap distribution F_n^*. Then, it is straightforward to show that, if F_n^* converges weakly to a sample-path continuous random cdf, then

$$P(\tau_n \leq q_n^*(1-\alpha)) = P(p_n^* \leq 1-\alpha) + o(1) \to 1-\alpha$$

This implies that a confidence interval of the form $[T_n - \rho(n)^{-1}q_n^*(1-\alpha), +\infty)$ has (unconditional) asymptotic coverage probability of $1-\alpha$. If the bootstrap is valid conditionally on some X_n, as in (3.2), then the (asymptotic) coverage is $1-\alpha$ also conditionally on this X_n. □

Our main results make extensive use of *joint* weak convergence in distribution. Should the notation not be self-explanatory, we refer to Appendix A for the formal definitions.

3.2 Unconditional bootstrap validity

The unconditional validity results in this section have in common the requirement, explicit or implicit, that the unconditional limit distribution of τ_n should be an average of the random limit distribution of τ_n^* given the data. Applications of Theorem 3.1 do not require a conditional analysis of τ_n, in contrast to applications of Theorem 3.2.

Theorem 3.1 *Let there exist a rv τ and a random element X, both defined on the same probability space, such that $(\tau_n, F_n^*) \xrightarrow{w} (\tau, F)$ in $\mathbb{R} \times \mathscr{D}(\mathbb{R})$ for $F_n^*(u) := P(\tau_n^* \leq u|D_n)$ and $F(u) := P(\tau \leq u|X)$, $u \in \mathbb{R}$. If the (possibly) random cdf F is sample-path continuous, then the bootstrap based on τ_n and τ_n^* is valid unconditionally.*

A trivial special case of Theorem 3.1 is obtained for independent τ and X. In this case the bootstrap cdf of τ_n^* estimates consistently the limiting unconditional cdf of τ_n and the bootstrap is valid in the usual sense.

In contrast, where the limit of the bootstrap cdf F_n^* is random (and even if it is continuous), the separate convergence facts $\tau_n \xrightarrow{w} \tau$ and $F_n^*(\cdot) \xrightarrow{w} F(\cdot) = P(\tau \leq \cdot|X)$ (or $\tau_n^* \xrightarrow{w}_w \tau|X$) are not sufficient for unconditional bootstrap validity. Some remarks on the joint convergence $(\tau_n, F_n^*) \xrightarrow{w} (\tau, F)$ are, hence, in order. A further strategy for proving it is outlined in Section 4.2.

Remark 3.4 An important special case of Theorem 3.1 involves stable convergence of the original statistic τ_n (see Häusler and Luschgy, 2015, p.33, for a definition). With the notation of Theorem 3.1, let the data D_n and the random element X be defined on the same probability space, whereas the rv τ be defined on an extension of this probability space. Assume that $\tau_n \to \tau$ $\sigma(X)$-stably and $F_n^* \xrightarrow{p} F := P(\tau \leq \cdot|X)$. Then $(\tau_n, F_n^*) \xrightarrow{w} (\tau, F)$ by Theorem 3.7(b) of Häusler and Luschgy (2015). For instance, in the statistical literature on integrated volatility, a result of the form $\tau_n \to \tau$ stably is contained in Theorem 3.1 of Jacod et al. (2009) for τ_n defined as a t-type statistic for integrated volatility, whereas the corresponding $F_n^* \xrightarrow{p} F$ result is established in Theorem 3.1 of Hounyo, Gonçalves and Meddahi (2017) for a combined wild and blocks-of-blocks bootstrap introduced in the latter paper.

Remark 3.5 More generally, if $\tau_n^* \xrightarrow{w^*}_w \tau^*|X$ and $(\tau_n^*, \tau_n, X_n) \xrightarrow{w} (\tau^*, \tau, X)$ for some D_n-measurable X_n $(n \in \mathbb{N})$, where the conditional distributions $\tau^*|X$ and $\tau|X$ are equal a.s. and have a sample-path continuous conditional cdf F, then $(\tau_n, F_n^*) \xrightarrow{w} (\tau, F)$; see Appendix B for additional details. $\square$

Alternatively, unconditional bootstrap validity could be established by means of an auxiliary conditional analysis of the original statistic τ_n. In the next theorem the conditioning sequence X_n is chosen such that the bootstrap statistic τ_n^* depends on the data D_n approximately through X_n (condition (†)). Then, the main requirement for bootstrap validity is that the limit bootstrap distribution should be a conditional average of the limit distribution of τ_n given X_n.

Theorem 3.2 *With the notation of Definition 1, let X_n be D_n-measurable ($n \in \mathbb{N}$). Let it hold that*

$$(P(\tau_n \leq \cdot | X_n), P(\tau_n^* \leq \cdot | D_n)) \xrightarrow{w} (F, F^*) \tag{3.3}$$

in $\mathscr{D}(\mathbb{R}) \times \mathscr{D}(\mathbb{R})$, where F and F^* are sample-path continuous random cdf's, and let

(†) there exist random elements X', X_n' such that F^* is X'-measurable, X_n' are X_n-measurable and $X_n' \xrightarrow{w} X'$ jointly with (3.3).

Then, if $E\{F(\cdot)|F^*\} = F^*(\cdot)$, the bootstrap based on τ_n and τ_n^* is valid unconditionally.

Remark 3.6 Under condition (†) of Theorem 3.2, $P(\tau_n^* \leq \cdot | X_n)$ and $P(\tau_n^* \leq \cdot | D_n)$ are both close to $P(\tau_n^* \leq \cdot | X_n')$, and in this sense τ_n^* depends on the data D_n approximately through X_n. Condition (†) is trivially satisfied in the case $F = F^*$ with the choice $X_n' = P(\tau_n \leq \cdot | X_n)$. It is also satisfied with $X_n' = P(\tilde{\tau}_n^* \leq \cdot | X_n)$ if $\tilde{\tau}_n^*$ is some measurable transformation of X_n and W_n^* such that $\tau_n^* = \tilde{\tau}_n^* + o_p(1)$ w.r.t. the probability measure on the space where D_n and W_n^* are jointly defined; see Appendix B. $\square$

Convergence (3.3) could be deduced from the weak convergence of the conditional distributions of τ_n and τ_n^*, as in the next corollary.

Corollary 3.1 *Let D_n and X_n ($n \in \mathbb{N}$) be as in Theorem 3.2. Let the rv τ and the random elements X, X' be defined on a single probability space and*

$$(\tau_n | X_n, \tau_n^* | D_n) \xrightarrow{w}_w (\tau | X, \tau | X') \tag{3.4}$$

in the sense of eq. (A.1). Let further $F(u) := P(\tau \leq u | X)$ and $F^*(u) := P(\tau \leq u | X')$, $u \in \mathbb{R}$, define sample-path continuous random cdf's. Then convergence (3.3) holds. Moreover, the bootstrap based on τ_n and τ_n^* is valid unconditionally provided that one of the following extra conditions holds:

(a) $X' = X$;

(b) $X = (X', X'')$ and $X_n' \xrightarrow{w} X'$ jointly with (3.4) for some X_n-measurable random elements X_n'.

Remark 3.7 An instance of (3.3) where F and F^* are not a.s. equal is provided by DGP (iii) of Section 2.2.3. There (3.4) holds with $\tau := M^{-1/2}(\omega_{\varepsilon|\eta}^{1/2}\xi_1 + (1 - \omega_{\varepsilon|\eta})^{1/2}\xi_2)$ and $X = (X', X'') = (M, (1 - \omega_{\varepsilon|\eta})^{1/2}\xi_2)$. Moreover, (3.4) is joint with the convergence $X_n' \xrightarrow{w} X'$ for $X_n' = n^{-2}M_n$ (see Appendix B). Hence, Corollary 3.1(b) implies that the bootstrap is unconditionally valid, as was already concluded in Section 2.2.1. $\square$

3.3 Conditional bootstrap validity

Theorem 3.3 below states the asymptotic behavior of the bootstrap p-value conditional on an X_n chosen to satisfy condition (†) of Theorem 3.2. It also characterizes the cases where the bootstrap is valid conditionally on such an X_n. Should validity conditional on such an X_n fail, in Corollary 3.2(b) we provide a result for validity conditional on a transformation of it.

Theorem 3.3 *Under the conditions of Theorem 3.2, the bootstrap p-value* p_n^* *satisfies*

$$P\left(p_n^* \leq q|\, X_n\right) \xrightarrow{w} F(F^{*-1}(q)) \tag{3.5}$$

for almost all $q \in (0,1)$*, and the bootstrap based on* τ_n *and* τ_n^* *is valid conditionally on* X_n *if and only if* $F = F^*$*, such that*

$$\sup_{u\in\mathbb{R}} \left|P\left(\tau_n \leq u|\, X_n\right) - P\left(\tau_n^* \leq u|\, D_n\right)\right| \xrightarrow{p} 0. \tag{3.6}$$

Remark 3.8 Convergence (3.6) means that the bootstrap distribution of τ_n^* consistently estimates the limit of the conditional distribution of τ_n given X_n. Although under condition (†) the proximity of $P\left(\tau_n \leq \cdot|\, X_n\right)$ and $P\left(\tau_n^* \leq \cdot|\, D_n\right)$ is necessary for bootstrap validity conditional on X_n, no such proximity is necessary for conditional validity in the general case. In fact, validity conditional on some X_n implies validity conditional on any measurable transformation $X_n' = \psi_n(X_n)$ and an analogue of (3.6) with X_n' in place of X_n cannot generally hold for all ψ_n, unless F^* is non-random. This is similar to what happens with unconditional bootstrap validity which, according to Theorem 3.1, may occur even if $P\left(\tau_n \leq \cdot\right)$ and $P\left(\tau_n^* \leq \cdot|\, D_n\right)$ are not close to each other. □

A corollary in the terms of weak convergence in distribution is given next.

Corollary 3.2 *Let* D_n, X_n *(*$n \in \mathbb{N}$*),* τ, F, F^* *be as in Corollary 3.1. Let (3.4) hold and* F, F^* *be sample-path continuous random cdf's. Then:*

(a) If $X' = X$*, the bootstrap based on* τ_n *and* τ_n^* *is valid conditionally on* X_n *and (3.6) holds.*

(b) If $X = (X', X'')$*,* $(X_n', X_n'') \xrightarrow{w} (X', X'')$ *jointly with (3.4) for some* X_n*-measurable random elements* (X_n', X_n'')*, and* $X_n''|X_n' \xrightarrow{w}_w X''|X'$*, then the bootstrap is valid conditionally on* X_n' *and (3.6) holds with* X_n *replaced by* X_n'*.*

Remark 3.9 Consider the linear regression example under the extra assumptions of Section 2.2.2 and set $\tau = (\int B_\eta^2)^{-1} \int B_\eta dB_\varepsilon$, $X = M$. It then follows (by using Theorem 3 of Georgiev *et al.*, 2019) that condition (3.4) holds in the form

$$(\tau_n|X_n, \tau_n^*|D_n) \xrightarrow{w}_w (\tau|B_\eta, \tau|B_\eta) = (1,1)N(0, \omega_\varepsilon M^{-1})|M \text{ a.s.}, \tag{3.7}$$

where $X_n := \{x_t\}_{t=1}^n$; equivalently, (3.3) holds with $F = F^* = \Phi(\omega_\varepsilon^{-1/2} M^{1/2}(\cdot))$. Hence, the bootstrap is consistent for the limit distribution of τ_n conditional on the regressor and, by Corollary 3.2(a), the bootstrap is valid conditionally on the regressor.

Remark 3.10 For DGP (iii) of Section 2.2.3, (3.4) holds with τ and $X = (X', X'')$ given in Remark 3.7. Moreover, (3.4) is joint with the convergence $(X_n', X_n'') \xrightarrow{w} (X', X'')$ for $X_n' = n^{-2}M_n$ and $X_n'' = M_n^{-1/2}\sum_{t=1}^{n} x_t E(\varepsilon_t|\eta_t)$ (see Appendix B). By Corollary 3.2(b), the bootstrap would be valid conditionally on M_n if it additionally holds that $X_n''|M_n \xrightarrow{w}_w (1-\omega_{\varepsilon|\eta})^{1/2}\xi_2|M = N(0, 1-\omega_{\varepsilon|\eta})$ a.s. □

Strategies for checking the convergence in (3.4) are outlined in Sections 4.2 and 4.3.

4 Applications

4.1 A permutation CUSUM test under infinite variance

Consider a standard CUSUM test for the null hypothesis (say, H_0) that $\{\varepsilon_t\}_{t=1}^{n}$ is a sequence of i.i.d. random variables. The test statistic is of the form

$$\tau_n := \nu_n^{-1} \max_{t=1,\ldots,n} \left|\sum\nolimits_{i=1}^{t} (\varepsilon_i - \bar{\varepsilon}_n)\right|, \; \bar{\varepsilon}_n := n^{-1}\sum\nolimits_{t=1}^{n} \varepsilon_t,$$

where ν_n is a permutation-invariant normalization sequence. Standard choices are $\nu_n^2 = \sum_{t=1}^{n}(\varepsilon_t - \bar{\varepsilon}_n)^2$ in the case where $E\varepsilon_t^2 < \infty$, and $\nu_n = \max_{t=1,\ldots,n}|\varepsilon_t|$ when $E\varepsilon_t^2 = \infty$. If ε_t is in the domain of attraction of a strictly α-stable law with $\alpha \in (0, 2)$, such that $E\varepsilon_t^2 = \infty$, the asymptotic distribution of τ_n depends on unknown parameters (e.g., the characteristic exponent α), which makes the test difficult to apply (see also Politis, Romano and Wolf, 1999, and the references therein). To overcome this problem, Aue et al. (2008) consider a permutation analogue of τ_n, defined as

$$\tau_n^* := \nu_n^{-1} \max_{t=1,\ldots,n} \left|\sum\nolimits_{i=1}^{t} (\varepsilon_{\pi(i)} - \bar{\varepsilon}_n)\right|$$

where π is a (uniformly distributed) random permutation of $\{1, 2, ..., n\}$, independent of the data.[3] In terms of Definition 1, the data is $D_n := \{\varepsilon_t\}_{t=1}^{n}$ and the auxiliary 'bootstrap' variate is $W_n^* := \pi$. With $X_n := \{\varepsilon_{(t)}\}_{t=1}^{n}$ denoting the vector of order statistics of $\{\varepsilon_t\}_{t=1}^{n}$, there exists a random permutation ϖ of $\{1, ..., n\}$ (under H_0, uniformly distributed conditionally on X_n) for which it holds that $\varepsilon_t = \varepsilon_{(\varpi(t))}$ $(t = 1, ..., n)$, whereas the 'bootstrap' sample is $\{\varepsilon_{\pi(t)}\}_{t=1}^{n}$. The results in Aue et al. (2008, Corollary 2.1, Theorem 2.4) imply that, if H_0 holds and ε_t is in the domain of attraction of a strictly α-stable law with $\alpha \in (0, 2)$, then $\tau_n \xrightarrow{w} \rho_\alpha(S)$ and $\tau_n^* \xrightarrow{w^*}_w \rho_\alpha(S)|S$ for a certain random function $\rho_\alpha(\cdot)$ and $S = (S_1, S_2)'$, with $S_i = \{S_{ij}\}_{j=1}^{\infty}$ $(i = 1, 2)$ being partial sums of sequences of i.i.d. standard exponential rv's, and with the function $\rho_\alpha(\cdot)$ independent of S.[4]

Aue et al. (2008) do not report the fact that inference is not invalidated by the failure of the permutation procedure to estimate consistently the distribution of $\rho_\alpha(S)$.

[3] The normalization of ν_n is only of theoretical importance for obtaining non-degenerate limit distributions. In practice, any bootstrap procedure comparing τ_n to the quantiles of τ_n^* is invariant to the choice of ν_n and can be implemented by setting $\nu_n = 1$.

[4] To avoid centering terms, Aue *et al.* (2008) assume additionally that the location parameter of the limit stable law is zero when $\alpha \in [1, 2)$. Moreover, although they provide conditional convergence results only for the finite-dimensional distributions of the CUSUM process, these could be strengthened to conditional functional convergence as in Proposition 1 of LePage *et al.* (1997) in order to obtain the conditional convergence of τ_n^*.

In fact, the situation is similar to that of Remark 2.1, as the conditional distributions $\tau_n|X_n$ and $\tau_n^*|D_n$ coincide a.s. under H_0. As a consequence, under H_0 the permutation test implements *exact*[5] finite-sample inference conditional on X_n and, additionally, the distribution of τ_n^* given the data estimates consistently the limit of the conditional distribution $\tau_n|X_n$, in the sense of joint weak convergence in distribution (see eq. (A.1)):

$$(\tau_n|X_n, \tau_n^*|D_n)' \xrightarrow{w}_w (\rho_\alpha(S)|S, \rho_\alpha(S)|S) \ . \tag{4.1}$$

CUSUM tests can also be applied to residuals from an estimated model in order to test for correct model specification or stability of the parameters (see e.g., Ploberger and Krämer, 1992). Consider thus the case where $\{\varepsilon_t\}_{t=1}^n$ are the disturbances in a statistical model (e.g., the regression model of Section 2), and we observe residuals $\hat{\varepsilon}_t$ obtained upon estimation of the model using a sample D_n not containing the unobservable $\{\varepsilon_t\}_{t=1}^n$. The residual-based CUSUM statistic is $\hat{\tau}_n := \hat{\nu}_n^{-1} \max_{t=1,\ldots,n} |\sum_{i=1}^t (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}_n)|$, where $\hat{\nu}_n$ and $\bar{\hat{\varepsilon}}_n$ are the analogues of ν_n and $\bar{\varepsilon}_n$ computed from $\hat{\varepsilon}_t$ instead of ε_t. The bootstrap statistic could be defined as $\hat{\tau}_n^* := \hat{\nu}_n^{-1} \max_{t=1,\ldots,n} |\sum_{i=1}^t (\hat{\varepsilon}_{\pi(i)} - \bar{\hat{\varepsilon}}_n)|$. If $\hat{\tau}_n - \tau_n \xrightarrow{p} 0$ and $(\hat{\tau}_n^* - \tau_n^*)|D_n \xrightarrow{w}_p 0$ under H_0 (e.g., due to consistent parameter estimation), then also $(\hat{\tau}_n - \tau_n)|X_n \xrightarrow{w}_p 0$, such that the (Lévy) distances between the pairs of conditional distributions $\hat{\tau}_n|X_n$ and $\tau_n|X_n$ on the one hand, and $\hat{\tau}_n^*|D_n$ and $\tau_n^*|D_n$ on the other hand, converge in probability to zero. Hence, in view of (4.1), and under the conjecture that $P(\rho_\alpha(S) \leq \cdot|S)$ defines a sample-path continuous cdf, the residual-based permutation procedure is consistent in the sense that

$$(\hat{\tau}_n|X_n, \hat{\tau}_n^*|D_n) \xrightarrow{w}_w (\rho_\alpha(S)|S, \rho_\alpha(S)|S) \tag{4.2}$$

for $X_n := \{\varepsilon_{(t)}\}_{t=1}^n$ again. It follows that: (i) the permutation residual-based test is valid conditionally on X_n, by Corollary 3.2(a) with condition (3.4) taking the form (4.2); (ii) this test is valid unconditionally, as a results of either the validity conditional on X_n, or by Corollary 3.1.

4.2 A parametric bootstrap goodness-of-fit test

The parametric bootstrap is a standard technique for the approximation of a conditional distribution of goodness-of-fit test statistics (Andrews, 1997; Lockhart, 2012). When these are discussed in the i.i.d. finite-variance setting, the limit of the bootstrap distribution is non-random. However, if we return to the relation (2.1), there exist relevant settings where a random limit of the normalized M_n implies that parametrically bootstrapped goodness-of-fit test statistics have random limit distributions.

4.2.1 Set up and a random limit bootstrap measure

Let the null hypothesis of interest, say H_0, be that the standardized errors $\omega_\varepsilon^{-1/2}\varepsilon_t$ in (2.1) have a certain known density f with mean 0 and variance 1. For expositional ease we assume that $\omega_\varepsilon = 1$ and is known to the econometrician. Then, the Kolmogorov-Smirnov statistic based on OLS residuals $\hat{\varepsilon}_t$ is

[5] By 'exact' we mean inference with respect to the true finite-sample (conditional) distribution of the test statistic.

$$\tau_n := n^{1/2} \sup_{s\in\mathbb{R}} \Big| n^{-1} \sum\nolimits_{t=1}^{n} \mathbb{I}_{\{\hat{\varepsilon}_t \le s\}} - \int_{-\infty}^{s} f \Big|.$$

A (parametric) bootstrap counterpart, τ_n^*, of τ_n could be constructed under H_0 by (i) drawing $\{\varepsilon_t^*\}_{t=1}^n$ as i.i.d. from f, independent of the data; (ii), regressing them on x_t, thus obtaining an estimator $\hat{\beta}^*$ and associated residuals $\hat{\varepsilon}_t^*$; and (iii) calculating τ_n^* as $\tau_n^* := n^{1/2} \sup_{s\in\mathbb{R}} |n^{-1} \sum_{t=1}^n \mathbb{I}_{\{\hat{\varepsilon}_t^* \le s\}} - \int_{-\infty}^s f|$.

To see that the distribution of the bootstrap statistic τ_n^* conditional on the data $D_n := \{x_t, y_t\}_{t=1}^n$ may have a random limit, consider the Gaussian case, $f = \Phi'$. Under the assumptions of Johansen and Nielsen (2016, Sec. 4.1-4.2), it holds (*ibidem*) that $\tau_n^* = \tilde{\tau}_n^* + o_p(1)$ under the product probability on the product probability space where the data and $\{\varepsilon_t^*\}$ are jointly defined, with

$$\tilde{\tau}_n^* := \sup_{s\in[0,1]} \Big| n^{-1/2} \sum_{t=1}^{n} (\mathbb{I}_{\{\varepsilon_t^* \le q(s)\}} - s) + \Phi'(q(s)) \hat{\beta}^* n^{-1/2} \sum_{t=1}^{n} x_t \Big|, \tag{4.3}$$

where $q(s) = \Phi^{-1}(s)$ is the s-th quantile of Φ. The expansion of τ_n^* holds also conditionally on the data, i.e., $\tau_n^* - \tilde{\tau}_n^* \xrightarrow{w^*}_p 0$, since convergence in probability to a constant is preserved upon such conditioning. Hence, if $\tilde{\tau}_n^* | D_n$ converges to a random limit, so does $\tau_n^* | D_n$ for the same limit. Assume that $X_n := n^{-\alpha/2} x_{\lfloor n \cdot \rfloor} \xrightarrow{w} U$ in $\mathscr{D}$ for some $\alpha > 0$ and that $M := \int U^2 > 0$ a.s. (e.g., $U = B_\eta$ if $x_t = \sum_{s=1}^{t-1} \eta_s$ with $\{\eta_t\}$ introduced in Section 2.2). Then $(M_n, \xi_n) := (\sum_{t=1}^n x_t^2, \sum_{t=1}^n x_t)$ satisfies $(n^{-\alpha-1} M_n, n^{-\alpha/2-1} \xi_n) \xrightarrow{w} (M, \xi)$, $\xi := \int U$. Furthermore, if $W_n^*(s) := n^{-1} \sum_{t=1}^n (\mathbb{I}_{\{\varepsilon_t^* \le q(s)\}} - s)$, $s \in [0,1]$, is the bootstrap empirical process in probability scale, then W_n^* and $M_n^{1/2} \hat{\beta}^*$ are independent of the data individually (the second one being conditionally standard Gaussian), but not jointly independent of the data, because

$$\mathrm{Cov}^*(n^{1/2} W_n^*(s), M_n^{1/2} \hat{\beta}^*) = (n^{-\alpha-1} M_n)^{-1/2} n^{-\alpha/2-1} \xi_n \psi(s) \xrightarrow{w} M^{-1/2} \xi \psi(s),$$

$s \in [0,1]$, where $\psi(\cdot) := E^*[\varepsilon_1^* \mathbb{I}_{\{\varepsilon_1^* \le q(\cdot)\}}] = -\Phi'(q(\cdot))$ is a trimmed mean function, with $\mathrm{Cov}^*(\cdot)$ and $E^*(\cdot)$ calculated under P^*. It is shown in Section S.4 that, more strongly,

$$(n^{1/2} W_n^*,\, n^{(\alpha+1)/2} \hat{\beta}^*,\, n^{-\alpha/2-1} \xi_n) \xrightarrow{w^*}_w (W, M^{-1/2} b, \xi) \big| (M, \xi) \tag{4.4}$$

on $\mathscr{D} \times \mathbb{R}^2$, where (W, b) is a pair of a standard Brownian bridge and a standard Gaussian rv individually independent of U (and thus, of M, ξ), but with Gaussian joint conditional (on U) distributions having covariance $\mathrm{Cov}(W(s), b|U) = M^{-1/2} \xi \psi(s)$, $s \in [0,1]$. Combining the expansion of τ_n^*, (4.3) and (4.4) with an extended CMT (Theorem A.1 in Appendix A) yields

$$\tau_n^* \xrightarrow{w^*}_w \Big\{ \sup_{s\in[0,1]} |W(s) + \Phi'(q(s)) M^{-1/2} b \xi| \Big\} \big| (M, \xi) = \tau \,|(M, \xi) \ \text{ a.s.}, \tag{4.5}$$

where $\tau := \sup_{s\in[0,1]} |\tilde{W}(s)|$ for a process $\tilde{W}$ which conditionally on U (and thus, on M, ξ), is a zero-mean Gaussian process with $\tilde{W}(0) = \tilde{W}(1) = 0$ a.s. and conditional covariance function $K(s, v) = s(1-v) - M^{-1} \xi^2 \psi(s) \psi(v)$ for $0 \le s \le v \le 1$. In summary, the limit bootstrap distribution is random because the latter conditional covariance is random whenever M or ξ are such.

4.2.2 Bootstrap validity

We now discuss in what sense τ_n^* can provide a distributional approximation of τ_n and whether the bootstrap can be valid in the sense of Definition 1.

Unconditional validity Under H_0 that $\varepsilon_t \sim$ i.i.d. $N(0,1)$, the bootstrap could be shown to be unconditionally valid using Theorem 3.1. Specifically, under H_0, the assumptions and results of Johansen and Nielsen (2016, Sec. 4.1-4.2) guarantee that τ_n has the expansion $\tau_n = \tilde{\tau}_n + o_p(1)$, with $\tilde{\tau}_n := \sup_{s\in[0,1]} |n^{-1/2}\sum_{t=1}^n(\mathbb{I}_{\{\varepsilon_t\leq q(s)\}} - s) + \Phi'(q(s))(\hat{\beta}-\beta)n^{-1/2}\sum_{t=1}^n x_t|$ defined similarly to $\tilde{\tau}_n^*$. Assume that $\hat{\beta}$ is asymptotically mixed Gaussian, such that jointly with $n^{-\alpha/2}x_{\lfloor n\cdot\rfloor} \overset{w}{\to} U$ it holds that

$$\Big(n^{-1/2}\sum_{t=1}^n(\mathbb{I}_{\{\varepsilon_t\leq q(s)\}} - s),\, n^{(\alpha+1)/2}(\hat{\beta}-\beta),\, n^{-\alpha/2-1}\xi_n\Big) \overset{w}{\to} (W, M^{-1/2}b, \xi) \ ;$$

then $\tau_n = \tilde{\tau}_n + o_p(1) \overset{w}{\to} \tau = \sup_{s\in[0,1]} |\tilde{W}(s)|$. Thus, the unconditional limit of τ_n obtains by averaging (over M,ξ) the conditional limit of τ_n^*. This is the main prerequisite for establishing unconditional bootstrap validity via Theorem 3.1. More precisely, it is proved in Section S.4 that

$$(\tau_n, F_n^*) \overset{w}{\to} (\tau, F)\,,\ F_n^*(\cdot) := P^*(\tau_n^* \leq \cdot),\ F(\cdot) := P(\tau \leq \cdot|M,\xi). \tag{4.6}$$

As F is sample-path continuous (e.g., by Proposition 3.2 of Linde, 1989, applied conditionally on M,ξ), Theorem 3.1 with $X := (M,\xi)$ guarantees unconditional validity.

Remark 4.1 We outline here our approach to the proof of (4.6), which is of interest also in other applications. The main ingredients are (a) the convergence $(\tau_n, X_n) \overset{w}{\to} (\tau, U)$; (b) the fact that its strong version $(\tau_n, X_n) \overset{a.s.}{\to} (\tau, U)$ can be shown to imply $\tau_n^* \overset{w^*}{\to}_p \tau|U$; and (c) the fact that the conditional distribution $\tau|U$, which is a.s. equal to $\tau|(M,\xi)$, is diffuse. The proof proceeds in two steps: (i) prove that $(\tau_n, X_n) \overset{w}{\to} (\tau, U)$; (ii) consider, by extended Skorokhod coupling (Corollary 5.12 of Kallenberg, 1997), a representation of D_n and (τ, U) such that, with an abuse of notation, $(\tau_n, X_n) \overset{a.s.}{\to} (\tau, U)$ and, on a product extension of the Skorokhod-representation space, prove that $\tau_n^* \overset{w^*}{\to}_p \tau|U$. The latter conditional assertion, due to the product structure of the probability space, can be proved as a collection of unconditional assertions by fixing the outcomes in the factor-space of the data. As F is sample-path continuous, $(\tau_n, F_n^*) \overset{p}{\to} (\tau, F)$ on the Skorokhod-representation space, whereas $(\tau_n, F_n^*) \overset{w}{\to} (\tau, F)$ on a general probability space. In other applications, the idea of a similar proof would be to choose X_n as D_n-measurable random elements such that τ_n^* depends on the data essentially through X_n.□

Conditional validity As $\tau_n = \tilde{\tau}_n + o_p(1)$ under H_0, with $\tilde{\tau}_n$ related to (M_n, ξ_n) through the same functional form as $\tilde{\tau}_n^*$, it is possible for $\tau_n|X_n$ to have the same random limit distribution under H_0 as τ_n^* given the data, i.e., $\tau_n|X_n \overset{w}{\to}_w \tau\,|(M,\xi)$. For instance, this occurs if $\{\varepsilon_t\}$ is an i.i.d. sequence independent of X_n, by the same argument as for $\tilde{\tau}_n^*$. Conditional validity can then be established through Corollary 3.2 by using the following fact.

Remark 4.2 The convergence in (3.4), required in Corollary 3.2, follows from the separate convergence facts $\tau_n|X_n \xrightarrow{w}_w \tau|X$, $\tau_n^* \xrightarrow{w^*}_w \tau^*|X'$ and $(\tau_n, \tau_n^*, \phi_n(X_n), \psi_n(D_n)) \xrightarrow{w} (\tau, \tau^*, X, X')$ for some measurable functions ϕ_n, ψ_n, provided that the conditional distributions $\tau|X'$ and $\tau^*|X'$ are equal a.s.; see Section S.4. □

Let $\phi_n(X_n) := \psi_n(X_n) := (n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n)$ and $X := X' := (M, \xi)$. By Remark 4.2, the convergence $\tau_n|X_n \xrightarrow{w}_w \tau|X'$, eq. (4.5) and the convergence $(\tau_n, \tau_n^*, \phi_n(X_n)) \xrightarrow{w} (\tau, \tau^*, X')$ with the distributions $\tau^*|X'$ and $\tau|X'$ equal a.s. (shown in the proof of (4.6), see Section S.4) are sufficient for eq. (3.4) to hold in the form

$$(\tau_n|X_n, \tau_n^*|D_n) \xrightarrow{w}_w (\tau|X', \tau|X').$$

As the random cdf F of the conditional distribution $\tau|X'$ is sample-path continuous, the bootstrap is valid conditionally on X_n by Corollary 3.2(a).

4.3 Bootstrap tests of parameter constancy

4.3.1 General set up

Here we apply the results of Section 3 to the classic problem of parameter constancy testing in regression models (see Andrews, 1993, and the references therein). Specifically, we deal with bootstrap implementations when the moments of the regressors may be unstable over time; see e.g. Hansen (2000) and Zhang and Wu (2012).

Consider a linear regression model for $y_{nt} \in \mathbb{R}$ given $x_{nt} \in \mathbb{R}^m$, in triangular array notation:

$$y_{nt} = \beta_t' x_{nt} + \varepsilon_{nt} \qquad (t = 1, 2, ..., n). \tag{4.7}$$

The null hypothesis of parameter constancy is $\mathsf{H}_0 : \beta_t = \beta_1$ $(t = 2, ..., n)$, which is tested here against the alternative $\mathsf{H}_1 : \beta_t = \beta_1 + \theta \mathbb{I}_{\{t \geq n^\star\}}$ $(t = 2, ..., n)$, where $n^\star := \lfloor r^\star n \rfloor$ and $\theta \neq 0$ respectively denote the timing and the magnitude of the possible break,[6] both assumed unknown to the econometrician. The so-called break fraction $r^\star$ belongs to a known closed interval $[\underline{r}, \overline{r}]$ in $(0, 1)$. In order to test H_0 against H_1, it is customary to consider the 'sup F' (or 'sup Wald') test (Andrews, 1993), based on the statistic $\mathscr{F}_n := \max_{r \in [\underline{r}, \overline{r}]} F_{\lfloor nr \rfloor}$, where $F_{\lfloor nr \rfloor}$ is the usual F statistic for testing the auxiliary null hypothesis that $\theta = 0$ in the regression

$$y_{nt} = \beta' x_{nt} + \theta' x_{nt} \mathbb{I}_{\{t \geq \lfloor rn \rfloor\}} + \varepsilon_{nt}.$$

We make the following assumption, allowing for non-stationarity in the regressors (see also Hansen, 2000, Assumptions 1 and 2).

Assumption $\mathcal{H}$. The following conditions on $\{x_{nt}, \varepsilon_{nt}\}$ hold:

(i) *(mda)* ε_{nt} *is a martingale difference array with respect to the current value of* x_{nt} *and the lagged values of* $(x_{nt}, \varepsilon_{nt})$*;*

(ii) *(wlln)* ε_{nt}^2 *satisfies the law of large numbers* $n^{-1} \sum_{t=1}^{\lfloor nr \rfloor} \varepsilon_{nt}^2 \xrightarrow{p} r(E\varepsilon_{nt}^2) = r\sigma^2 > 0$, *for all* $r \in (0, 1]$*;*

[6] We suppress the possible dependence of $\beta_t = \beta_{nt}$ on n with no risk of ambiguities.

(iii) *(non-stationarity)* $(M_n, V_n, N_n) \xrightarrow{w} (M, V, N)$ *in* $\mathscr{D}_{m\times m} \times \mathscr{D}_{m\times m} \times \mathscr{D}_m$ for

$$(M_n, V_n, N_n) := \left(\tfrac{1}{n}\sum\nolimits_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}', \tfrac{1}{n\sigma^2}\sum\nolimits_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}'\varepsilon_{nt}^2, \tfrac{1}{n^{1/2}\sigma}\sum\nolimits_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}\varepsilon_{nt}\right)$$

and where M *and* V *are a.s. continuous and (except at 0) strictly positive-definite valued processes, whereas* N*, conditionally on* $\{V, M\}$*, is a zero-mean Gaussian process with covariance kernel* $E\{N(r_1)\,N(r_2)'\} = V(r_1)$ $(0 \le r_1 \le r_2 \le 1)$.

Remark 4.3 A special case of Assumption $\mathcal{H}$ is obtained when the regressors satisfy the weak convergence $x_{n\lfloor n\cdot\rfloor} \xrightarrow{w} U$ in $\mathscr{D}_m$, such that $M = \int_0^{\cdot} UU'$. Under extra conditions (e.g., if $\sup_n \sup_{t=1,\ldots,n} E|E(\varepsilon_{nt}^2 - \sigma^2|\mathcal{F}_{n,t-i})| \to 0$ as $i \to \infty$ for some filtrations $\mathcal{F}_{n,t}$, $n \in \mathbb{N}$, to which $\{\varepsilon_{nt}^2\}$ is adapted), also $V = \int_0^{\cdot} UU'$ (see Theorem A.1 of Cavaliere and Taylor, 2009). □

The null asymptotic distribution of $\mathscr{F}_n$ under Assumption $\mathcal{H}$ is provided in Hansen (2000, Theorem 2):

$$\mathscr{F}_n \xrightarrow{w} \sup_{r\in[\underline{r},\bar{r}]} \{\tilde{N}(r)'\tilde{M}(r)^{-1}\tilde{N}(r)\} \tag{4.8}$$

with $\tilde{N}(u) := N(u) - M(u)\,M(1)^{-1}N(1)$ and $\tilde{M}(r) := M(r) - M(r)\,M(1)^{-1}M(r)$. In the case of (asymptotically) stationary regressors, $\mathscr{F}_n$ converges to the supremum of a squared tied-down Bessel process; see Andrews (1993). In the general case, however, since the asymptotic distribution in (4.8) depends on the joint distribution of the limiting processes M, N, V, which is unspecified under Assumption $\mathcal{H}$, asymptotic inference based on (4.8) is unfeasible. Simulation methods as the bootstrap can therefore be appealing devices for computing p-values associated with $\mathscr{F}_n$.

4.3.2 Bootstrap test and random limit bootstrap distribution

Following Hansen (2000), we consider here a fixed-regressor wild bootstrap intended to accommodate possible conditional heteroskedasticity of ε_{nt}. It is based on the residuals $\tilde{e}_{nt}$ from the OLS regression of y_{nt} on x_{nt} and $x_{nt}\mathbb{I}_{\{t\ge\lfloor\tilde{r}n\rfloor\}}$, where $\tilde{r} := \arg\max_{r\in[\underline{r},\bar{r}]} F_{\lfloor nr\rfloor}$ is the estimated break fraction for the original sample. The bootstrap statistic is $\mathscr{F}_n^* := \max_{r\in[\underline{r},\bar{r}]} F_{\lfloor nr\rfloor}^*$, where $F_{\lfloor nr\rfloor}^*$ is the F statistic for $\theta^* = 0$ in the auxiliary regression

$$y_t^* = \beta^{*\prime}x_{nt} + \theta^{*\prime}x_{nt}\mathbb{I}_{\{t\ge\lfloor rn\rfloor\}} + \text{error}_{nt}^*, \tag{4.9}$$

with bootstrap data $y_t^* := \tilde{e}_{nt}w_t^*$ for an i.i.d. $N(0,1)$ sequence of bootstrap multipliers w_t^* independent of the data (as in Hansen, 2000, we set without loss of generality $\beta = 0$ in the bootstrap sample). The weak limit of $\mathscr{F}_n^*$ given the data is stated next.

Theorem 4.1 *Under Assumption* $\mathcal{H}$ *and under* H_0 *it holds that, with* $\tilde{M}, \tilde{N}$ *as in (4.8),*

$$\mathscr{F}_n^* \xrightarrow{w^*}_w \sup_{r\in[\underline{r},\bar{r}]} \{\tilde{N}(r)'\tilde{M}(r)^{-1}\tilde{N}(r)\}\big|(M, V)\ . \tag{4.10}$$

Remark 4.4 Theorem 4.1 establishes that, in general, the limit distribution of the fixed-regressor bootstrap statistic is *random.* In particular, it is distinct from the limit in eq. (4.8) and, as a result, the bootstrap does not estimate consistently the unconditional limit distribution of the statistic $\mathscr{F}_n$ under H_0 (contrary to the claim in Theorem 6 of Hansen, 2000). To illustrate the limiting randomness, consider the case $M = V$ with a scalar regressor $x_{nt} \in \mathbb{R}$. By a change of variable (as in Theorem 3 of Hansen, 2000), convergence (4.10) reduces to

$$\mathscr{F}_n^* \overset{w^*}{\to}_w \sup_{u \in I(M,\underline{r},\bar{r})} \left\{ \tfrac{W(u)^2}{u(1-u)} \right\} \Big| M \quad \text{for} \quad I(M,\underline{r},\bar{r}) := \left[\tfrac{M(\underline{r})}{M(1)}, \tfrac{M(\bar{r})}{M(1)} \right],$$

where W is a standard Brownian bridge on $[0,1]$, independent of M. As the maximization interval $I(M,\underline{r},\bar{r})$ depends on M, so does the supremum itself. □

4.3.3 Bootstrap validity

Although under Assumption $\mathcal{H}$ the bootstrap does not replicate the asymptotic (unconditional) distribution in (4.8), unconditional bootstrap validity can be established under no further assumptions than Assumption $\mathcal{H}$, by using the results in Section 3.2. In contrast, if interest is in achieving bootstrap validity conditional on the regressors $X_n := \{x_{nt}\}_{t=1}^n$, as it may appear natural when the regressors are kept fixed across bootstrap samples, further conditions are required; e.g., the following Assumption $\mathcal{C}$.

Assumption $\mathcal{C}$. *Assumption $\mathcal{H}$ holds and, jointly with the convergence in Assumption $\mathcal{H}$*(iii)*, it holds that $(M_n, V_n, N_n)| X_n \overset{w}{\to}_w (M, V, N)\,|(M, V)$ as random measures on $\mathscr{D}_{m\times m} \times \mathscr{D}_{m\times m} \times \mathscr{D}_m$.*

Remark 4.5 Assumption $\mathcal{C}$ is stronger than Assumption $\mathcal{H}$ due to the fact that, differently from the bootstrap variates w_t^*, the errors $\{\varepsilon_{nt}\}$ need not be independent of $\{x_{nt}\}$. The third DGP of Section 2.2.3 could be used to construct an example, with $x_{nt} := n^{-1/2}x_t$ and $\varepsilon_{nt} := \varepsilon_t$, where Assumption $\mathcal{H}$(iii) holds but Assumption $\mathcal{C}$ does not.

Remark 4.6 The meaning of 'jointly' in Assumption $\mathcal{C}$ is given in eq. (A.2). By Lemma A.1(b), the convergence in Assumption $\mathcal{C}$ will be joint with that in Assumption $\mathcal{H}$(iii) if $n^{-1}\sigma^{-2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}'(\varepsilon_{nt}^2 - E(\varepsilon_{nt}^2|X_n)) = o_p(1)$ in $\mathscr{D}_{m\times m}$, such that the process $n^{-1}\sigma^{-2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}'\varepsilon_{nt}^2$ is asymptotically equivalent to an X_n-measurable process. □

The results on the validity of the bootstrap parameter constancy tests are summarized in the following theorem.

Theorem 4.2 *Let the parameter constancy hypothesis H_0 hold for model (4.7). Then, under Assumption $\mathcal{H}$, the bootstrap based on $\tau_n = \mathscr{F}_n$ and $\tau_n^* = \mathscr{F}_n^*$ is unconditionally valid. If Assumption $\mathcal{C}$ holds, then the bootstrap based on $\mathscr{F}_n$ and $\mathscr{F}_n^*$ is valid also conditionally on X_n.*

Theorem 4.2 under Assumption $\mathcal{H}$ is proved along the lines of Remark 4.1. Under Assumption $\mathcal{C}$ the proof could be recast in terms of the following general strategy to check condition (3.4) of Corollary 3.2(a), with $\phi_n(X_n) := \psi_n(D_n) := (M_n, V_n)$ and $X := X' := (M, V)$.

Remark 4.7 With the notation of Remark 4.2, convergence (3.4) follows from $\tau_n|X_n \overset{w}{\to}_w \tau|X$ and $(\tau_n, \phi_n(X_n), \psi_n(D_n)) \overset{w}{\to} (\tau, X, X')$ together with the implication (when it holds) from $\psi_n(D_n) \overset{a.s.}{\to} X'$ to $\tau_n^* \overset{w^*}{\to}_p \tau^*|X'$, provided that the conditional distributions $\tau|X'$ and $\tau^*|X'$ are a.s. equal. The convergence $\tau_n|X_n \overset{w}{\to}_w \tau|X$ is the new ingredient compared to Remark 4.1. An implementation strategy is: (i) prove that $\tau_n|X_n \overset{w}{\to}_w \tau|X$ and $(\tau_n, \phi_n(X_n), \psi_n(D_n)) \overset{w}{\to} (\tau, X, X')$; (ii) consider a Skorokhod representation of D_n and (τ, X, X') such that, maintaining the notation, $(\tau_n, \phi_n(X_n), \psi_n(D_n)) \overset{a.s.}{\to} (\tau, X, X')$ and, as a result, $\tau_n|X_n \overset{w}{\to}_w \tau|X$ strengthens to $\tau_n|X_n \overset{w}{\to}_p \tau|X$ (see Lemma A.1 in Appendix A); (iii) redefine the bootstrap variates W_n^* on a product extension of the Skorokhod-representation space and prove there that $\tau_n^* \overset{w^*}{\to}_p \tau^*|X'$. Then (3.4) holds on a general probability space. Notice also that if $\phi_n(X_n) = (X_n', X_n'')$ and $\psi_n(D_n) = X_n'$, then the convergence $(X_n', X_n'') \overset{w}{\to} (X', X'')$ in Corollary 3.2(b) is joint with (3.4).□

5 Conclusions

When the distribution of a bootstrap statistic conditional on the data is random in the limit, the bootstrap fails to estimate consistently the asymptotic distribution of the original statistic. Renormalization of the statistic of interest cannot always be used as a way to eliminate the limiting bootstrap randomness (e.g., it cannot be used in any of the applications in Section 4). Nevertheless, we have shown that if bootstrap validity is defined as (large sample) control over the frequency of correct inferences, then randomness of the limit bootstrap distribution does not imply invalidity of the bootstrap, even without renormalizing the original statistic. A bootstrap scheme, therefore, need not be discarded for the sole reason of giving rise to a random limit bootstrap measure.

For the asymptotic validity of bootstrap inference, in an unconditional or a conditional sense, we have established sufficient conditions and strategies to verify these conditions in specific applications. The conditions differ mainly in their demands on the dependence structure of the data, and are more restrictive for conditional validity to hold.

We have provided three applications to well-known econometric inference problems which feature randomness of the limit bootstrap distribution. As usual, alternative bootstrap schemes giving rise to non random bootstrap measures could also be put forward and in practice the choice of a bootstrap scheme to use would have to be made on a case-by-case basis. For instance, in the CUSUM application of Section 4.1, the m out of n bootstrap could be consistent for the unconditional asymptotic distribution of the statistic of interest; however, differently to the permutation test of Aue et al. (2008), which gives rise to a random limit measure, it would not be *exact* in finite samples. For the structural break test of Section 4.3, the use of the fixed regressor bootstrap (and, consequently, the randomness of the limit bootstrap measure) cannot be avoided, given the level of generality assumed on the regressors and on the error terms. However, if the regressors were known to be I(1), then a recursive bootstrap (with unit roots imposed) could in principle be implemented and it would mimic the unconditional limit distribution of the statistic of interest, rather than a conditional limit. Which of the two bootstraps would be preferable in terms of size and power requires additional investigation.

Among the further applications that could be analyzed using our approach are bootstrap inference in weakly or partially identified models, inference in time series models with time-varying (stochastic) volatility, inference after model selection, and the bootstrap in high-dimensional models. In addition, the methods we propose could be useful in problems involving nuisance parameters that are not consistently estimable under the null hypothesis but where sufficient statistics are available (with the bootstrap being potentially valid conditionally on such statistics). In these problems it could be of further interest to study conservative inferential procedures satisfying a weaker condition than (3.1), e.g., $\liminf P(p_n^* \geq q) \geq 1 - q$ for all $q \in (0, 1)$ in the case of tests rejecting for small values of p_n^*.

An important issue not analyzed in the paper is whether the bootstrap can deliver refinements over standard asymptotics in cases where the limit bootstrap measure is random. We have seen in Sections 2 and 4.1 that bootstrap inference in such cases could be exact or close to exact. This seems to suggest that a potential for refinements exists. Moreover, there is also a potential for the bootstrap to inherit the finite-sample refinements offered by conditional asymptotic expansions (in line with Barndorff-Nielsen's p^*-formula, see Barndorff-Nielsen and Cox, 1994, Sec. 6.2), as has been established for some bootstrap procedures (DiCiccio and Young, 2008) in the special case of correctly specified parametric models. The study of such questions requires mathematical tools different from those employed here and is left for further research.

References

Andrews, D.W.K. (1993): Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821–856.

—— (1997): A conditional Kolmogorov test, *Econometrica* 65, 1097–1128.

—— (2000): Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space, *Econometrica*, 68, 399–405.

Athreya, K.B.(1987): Bootstrap of the mean in the infinite variance case, *The Annals of Statistics* 15, 724-731.

Aue, A., I. Berkes and L. Horvàth (2008): Selection from a stable box,*Bernoulli* 14, 125–139.

Barndorff-Nielsen, O.E. and D.R. Cox (1994): *Inference and Asymptotics*, Chapman & Hall.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, and R.L. Taylor (1991): Bootstrapping unstable first-order autoregressive processes, *The Annals of Statistics* 19, 1098–1101.

Beran, R. (1997): Diagnosing Bootstrap success, *Annals of the Institute of Statistical Mathematics* 49, 1–24.

Billingsley, P. (1968): *Convergence of Probability Measures*, John Wiley: New York.

Cattaneo, M., M. Jansson and K. Nagasawa (2020): Bootstrap-based inference for cube root asymptotics, *Econometrica*, forthcoming.

Cavaliere, G. and I. Georgiev (2020): Inference under random limit bootstrap measures: supplemental material.

Cavaliere, G., I. Georgiev and A.M.R. Taylor (2013): Wild bootstrap of the mean in the infinite variance case, *Econometric Reviews* 32, 204–219.

Cavaliere, G., H.B. Nielsen and A. Rahbek (2015): Bootstrap testing of hypotheses on co-integration relations in vector autoregressive models, *Econometrica*, 83, 813–831.

Cavaliere, G. and A.M.R. Taylor (2009): Heteroskedastic time series with a unit root, *Econometric Theory* 25, 1228–1276.

Chan, N.H. and C.Z. Wei (1988): Limiting distributions of Least Squares estimates of unstable autoregressive processes, *Annals of Statistics* 16, 367–401.

Crimaldi, I. and L. Pratelli (2005): Convergence results for conditional expectations, *Bernoulli* 11, 737–745.

DasGupta, A. (2008): *Asymptotic Theory of Statistics and Probability*, Springer-Verlag: Berlin.

DiCiccio, T. and G.A. Young (2008): Conditional properties of unconditional parametric bootstrap procedures for inference in exponential families, *Biometrika* 95, 747–758.

Georgiev, I., D. Harvey, S. Leybourne and A.M.R. Taylor (2019): A bootstrap stationarity test for predictive regression invalidity, *Journal of Business & Economic Statistics* 37, 528–541.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, Springer-Verlag: Berlin.

Hansen, B.E. (2000): Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.

Häusler, E. and H. Luschgy (2015): *Stable Convergence and Stable Limit Theorems*, Springer-Verlag: Berlin.

Horowitz, J.L. (2001): The bootstrap. In Heckman, J.J. and E. Leamer (eds.) *Handbook of Econometrics* 5, chapter 52, Elsevier: Amsterdam.

Hounyo, U., S. Gonçalves and N. Meddahi (2017): Bootstrapping pre-averaged realized volatility under market microstructure noise, *Econometric Theory* 33, 791–838.

Jacod, J., Y. Li, P. Mykland, M. Podolskij and M. Vetter (2009): Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications* 119, 2249-2276.

Johansen, S. and B. Nielsen (2016): Analysis of the Forward Search using some new results for martingales and empirical processes, *Bernoulli* 22, 1131–1183.

Kallenberg, O. (1997): *Foundations of Modern Probability*, Springer-Verlag: Berlin.

—— (2017): *Random measures: theory and applications*, Springer-Verlag: Berlin.

Knight, K. (1989): On the bootstrap of the sample mean in the infinite variance case, *The Annals of Statistics* 17, 1168-1175.

Lahiri, S.N. (1996): Effects of block lengths on the validity of block resampling methods, *Probability Theory & Related Fields* 121, 73–97.

LePage, R. and K. Podgórsky (1996): Resampling permutations in regression without second moments, *Journal of Multivariate Analysis* 57, 119–141.

LePage, R., K. Podgórsky and M. Ryznar (1997): Strong and conditional invariance principles for samples attracted to stable laws, *Probability Theory and Related Fields* 108, 281–298.

Linde, W. (1989): Gaussian measure of translated balls in a Banach space, *Theory of Probability and its Applications* 34, 349–359.

Lockhart, R. (2012): Conditional limit laws for goodness-of-fit tests, *Bernoulli* 18, 857–882.

Politis, D., J. Romano and M. Wolf (1999): *Subsampling*, Springer, Berlin.

Ploberger, W. and W. Krämer (1992): The CUSUM test with OLS residuals, *Econometrica* 60, 271–285.

Reid, N. (1995): The roles of conditioning in inference, *Statistical Science* 10, 138–199.

Rubshtein, B. (1996): A central limit theorem for conditional distributions, In Bergelson V., P. March, J. Rosenblatt (eds.), *Convergence in Ergodic Theory and Probability*, De Gruyter: Berlin.

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

Shao, X. and D.N. Politis (2013): Fixed b subsampling and the block bootstrap: improved confidence sets based on p-value calibration, *Journal of the Royal Statistical Society B* 75, 161–184.

Sweeting, T.J. (1989): On conditional weak convergence, *Journal of Theoretical Probability 2*, 461–474.

Zhang, T. and W. Wu (2012): Inference of time-varying regression models, *Annals of Statistics* 40, 1376–1402.

Appendices

A Weak convergence in distribution

In this section we establish some properties of weak convergence in distribution for random elements of Polish spaces. They are useful in applications, in order to verify the high-level conditions of our main theorems, as well as to prove these very theorems. Recall the convention that, throughout, Polish spaces are equipped with their Borel sets. Finite k-tuples of random elements defined on the same probability space are considered as random elements of a product space with the product topology and σ-algebra.

Let (Z_n, X_n) and (Z, X) be random elements such that $Z_n = (Z'_n, Z''_n)$ and $Z = (Z', Z'')$ are $\mathcal{S}'_Z \times \mathcal{S}''_Z$-valued, whereas $X_n = (X'_n, X''_n)$ and $X = (X', X'')$ are resp. $\mathcal{S}' \times \mathcal{S}''$-valued and $\mathcal{S}'_X \times \mathcal{S}''_X$-valued ($n \in \mathbb{N}$). We say that $Z'_n|X'_n \xrightarrow{w}_w Z'|X'$ and $Z''_n|X''_n \xrightarrow{w}_w Z''|X''$ *jointly* (denoted by $(Z'_n|X'_n, Z''_n|X''_n) \xrightarrow{w}_w (Z'|X', Z''|X'')$) if

$$\big(E\{h'(Z'_n)|X'_n\},\, E\{h''(Z''_n)|X''_n\}\big) \xrightarrow{w} \big(E\{h'(Z')|X'\},\, E\{h''(Z'')|X''\}\big) \tag{A.1}$$

for all $h' \in \mathcal{C}_b(\mathcal{S}'_Z)$ and $h'' \in \mathcal{C}_b(\mathcal{S}''_Z)$. Even for $X'_n = X''_n$, this property is weaker than the convergence $(Z'_n, Z''_n)|X'_n \xrightarrow{w}_w (Z', Z'')|X$ defined by $E\{g(Z'_n, Z''_n)|X'_n\} \xrightarrow{w} E\{g(Z', Z'')|X\}$ for all $g \in \mathcal{C}_b(\mathcal{S}'_Z \times \mathcal{S}''_Z)$. We notice that for $Z'_n = X'_n$, (A.1) reduces to

$$\big(Z'_n, E\{h''(Z''_n)|X''_n\}\big) \xrightarrow{w} \big(Z', E\{h''(Z'')|X''\}\big) \tag{A.2}$$

for all $h'' \in \mathcal{C}_b(\mathcal{S}''_Z)$ and in this case we write $(Z'_n, (Z''_n|X''_n)) \xrightarrow{w}_w (Z'', (Z'|X'))$ (see Corollary S.1 in Section S.2).

The first lemma given here is divided in two parts. In the first part, we provide conditions for strengthening weak convergence in distribution to weak convergence in probability. The second part, in its simplest form, provides conditions such that the two convergence facts $(Z_n, X_n) \xrightarrow{w} (Z, X)$ and $Z_n|X_n \xrightarrow{w}_w Z|X$ imply the joint convergence $((Z_n|X_n), Z_n, X_n) \xrightarrow{w}_w ((Z|X), Z, X)$.

Lemma A.1 *Let $\mathcal{S}_Z$, $\mathcal{S}'_Z, \mathcal{S}_X$ and $\mathcal{S}'_X$ be Polish spaces. Consider the random elements Z_n, Z ($\mathcal{S}_Z$-valued), Z'_n, Z' ($\mathcal{S}'_Z$-valued), X_n ($\mathcal{S}_X$-valued) and X'_n, X ($\mathcal{S}'_X$-valued) for $n \in \mathbb{N}$. Assume that X'_n are X_n-measurable and $Z_n|X_n \xrightarrow{w}_w Z|X$.*

(a) If all the considered random elements are defined on the same probability space, $(Z_n, X'_n) \xrightarrow{w} (Z, X)$ and $X'_n \xrightarrow{p} X$, then $Z_n|X_n \xrightarrow{w}_p Z|X$.

(b) If $(Z_n, X'_n, Z'_n) \xrightarrow{w} (Z, X, Z')$, then the joint convergence $((Z_n|X_n), Z_n, X'_n, Z'_n) \xrightarrow{w}_w ((Z|X), Z, X, Z')$ holds in the sense that, for all $h \in \mathcal{C}_b(\mathcal{S}_Z)$,

$$(E\{h(Z_n)|X_n\}, Z_n, X'_n, Z'_n) \xrightarrow{w} (E\{h(Z)|X\}, Z, X, Z'). \tag{A.3}$$

Notice that, by choosing $Z'_n = Z' = 1$, a corollary of Lemma A.1(b) not involving Z'_n and Z' is obtained. It states that $Z_n|X_n \xrightarrow{w}_w Z|X$ and $(Z_n, X'_n) \xrightarrow{w} (Z, X)$ together imply the joint convergence $((Z_n|X_n), Z_n, X'_n) \xrightarrow{w}_w ((Z|X), Z, X)$, provided that X'_n are X_n-measurable.

By means of eq. (A.1) we defined *joint* weak convergence in distribution and denoted it by $(Z'_n|X'_n, Z''_n|X''_n) \xrightarrow{w}_w (Z'|X', Z''|X'')$. We now extend it to

$$((Z'_n|X'_n), (Z''_n|X''_n), Z'''_n) \xrightarrow{w}_w ((Z'|X'), (Z''|X''), Z'''), \tag{A.4}$$

defined to mean that

$$(E\{h'(Z'_n)|X'_n\}, E\{h''(Z''_n)|X''_n\}, Z'''_n) \xrightarrow{w} (E\{h'(Z')|X'\}, E\{h''(Z'')|X''\}, Z''') \tag{A.5}$$

for all $h' \in \mathcal{C}_b(\mathcal{S}'_Z)$ and $h'' \in \mathcal{C}_b(\mathcal{S}''_Z)$. The natural equivalence of $((Z'_n|X'_n), (Z'_n|X'_n), Z'''_n)$ $\xrightarrow{w}_w ((Z'|X'), (Z'|X'), Z''')$ and $((Z'_n|X'_n), Z'''_n) \xrightarrow{w}_w ((Z'|X'), Z''')$ holds under separability of the space $\mathcal{S}'''$ where Z'''_n, Z''' take values (see Remark S.1).

In Lemma A.2(b) below we relate (A.4) to the joint weak convergence of the respective conditional cdf's in the case of rv's Z'_n, Z''_n, Z' and Z''. Before that, in Lemma A.2(a) we show how joint weak convergence can be strengthened to a.s. weak convergence on a special probability space. For a single convergence $Z'_n|X'_n \xrightarrow{w}_w Z'|X'$, part (a) implies that there exists a Skorokhod representation $(\tilde{Z}'_n, \tilde{X}'_n) \stackrel{d}{=} (Z'_n, X'_n)$, $(\tilde{Z}', \tilde{X}') \stackrel{d}{=} (Z', X')$ such that $\tilde{Z}'_n|\tilde{X}'_n \xrightarrow{w}_{a.s.} \tilde{Z}'|\tilde{X}'$.

LEMMA A.2 *Let $(Z'_n, Z''_n, Z'''_n, X'_n, X''_n)$ and (Z', Z'', Z''', X', X'') be random elements of the same Polish product space, defined on possibly different probability spaces ($n \in \mathbb{N}$).*

(a) If (A.4)-(A.5) hold, then there exist a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and random elements $(\tilde{X}'_n, \tilde{X}''_n, \tilde{Z}'_n, \tilde{Z}''_n, \tilde{Z}'''_n) \stackrel{d}{=} (X'_n, X''_n, Z'_n, Z''_n, Z'''_n)$, $(\tilde{X}', \tilde{X}'', \tilde{Z}', \tilde{Z}'', \tilde{Z}''') \stackrel{d}{=} (X', X'', Z', Z'', Z''')$ defined on this space such that $\tilde{Z}'_n|\tilde{X}'_n \xrightarrow{w}_{a.s.} \tilde{Z}'|\tilde{X}', \tilde{Z}''_n|\tilde{X}''_n \xrightarrow{w}_{a.s.} \tilde{Z}''|\tilde{X}''$ and $\tilde{Z}'''_n \xrightarrow{a.s.} \tilde{Z}'''$.

(b) Let Z', Z'' be rv's and Z''' be $\mathcal{S}'''$-valued. If the conditional distributions $Z'|X'$ and $Z''|X''$ are diffuse, then (A.4)-(A.5) is equivalent to the weak convergence of the associated random cdf's:

$$\big(P(Z'_n \leq \cdot|X'_n),\, P(Z''_n \leq \cdot|X''_n), Z'''_n\big) \xrightarrow{w} \big(P(Z' \leq \cdot|X'),\, P(Z'' \leq \cdot|X''), Z'''\big) \tag{A.6}$$

as random elements of $\mathscr{D}(\mathbb{R}) \times \mathscr{D}(\mathbb{R}) \times \mathcal{S}'''$.

The definition of the convergence $Z_n|X_n \xrightarrow{w}_w Z|X$ implies that $h(Z_n)|X_n \xrightarrow{w}_w h(Z)|X$ for any continuous $h: \mathcal{S}_Z \to \mathcal{S}'_Z$ between Polish spaces. A generalization for functions h with a negligible set of discontinuities is provided in the following CMT (for weak convergence a.s. and weak convergence in probability, see Theorem 10 of Sweeting, 1989).

THEOREM A.1 *Let $\mathcal{S}_Z, \mathcal{S}'_Z, \mathcal{S}_X$ and $\mathcal{S}'_X$ be Polish spaces and the random elements Z_n, Z be $\mathcal{S}_Z$-valued, X_n be $\mathcal{S}_X$-valued and X be $\mathcal{S}'_X$-valued. If $Z_n|X_n \xrightarrow{w}_w Z|X$ and $h: \mathcal{S}_Z \to \mathcal{S}'_Z$ has its set of discontinuity points D_h with $P(Z \in D_h|X) = 0$ a.s., then $h(Z_n)|X_n \xrightarrow{w}_w h(Z)|X$.*

Next, we prove in Theorem A.2 a weak convergence result for iterated conditional expectations. The theorem provides conditions under which the convergence $E(z_n|X_n) \xrightarrow{w} E(z|X', X'')$ implies, upon iteration of the expectations, that $E\{E(z_n|X_n)|X'_n\} \xrightarrow{w}$

$E\{E(z|X', X'')|X'\}$ for rv's z_n, z and for X_n-measurable X'_n. In terms of weak convergence in distribution, the result allows to pass from $Z_n|X_n \xrightarrow{w}_w Z|(X', X'')$ to $Z_n|X'_n \xrightarrow{w}_w Z|X'$. We need, however, a more elaborate version for joint weak convergence.

Theorem A.2 *For $n \in \mathbb{N}$, let z_n be integrable rv's, X_n, Y_n and (X'_n, X''_n) be random elements of Polish spaces (say, $\mathcal{S}_X$, $\mathcal{S}_Y$ and $\mathcal{S}'_X$), defined on the probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ and such that (X'_n, X''_n) are X_n-measurable ($n \in \mathbb{N}$). Let also z be an integrable rv and $Y, (X', X'')$ be random elements of the Polish spaces $\mathcal{S}_Y$, $\mathcal{S}'_X$ defined on a probability space $(\Omega, \mathcal{F}, P)$. If*

$$(E(z_n|X_n), X'_n, X''_n, Y_n) \xrightarrow{w} (E(z|X', X''), X', X'', Y) \tag{A.7}$$

and $X''_n|X'_n \xrightarrow{w}_w X''|X'$, then $E(z_n|X'_n) \xrightarrow{w} E(z|X')$ jointly with (A.7).

Moreover, let Z_n, Z be random elements of a Polish space $\mathcal{S}_Z$ defined resp. on $(\Omega_n, \mathcal{F}_n, P_n)$ and $(\Omega, \mathcal{F}, P)$. If

$$((Z_n|X_n), X'_n, X''_n, Y_n) \xrightarrow{w}_w ((Z|(X', X'')), X', X'', Y) \tag{A.8}$$

and $X''_n|X'_n \xrightarrow{w}_w X''|X'$, then

$$((Z_n|X'_n), (Z_n|X_n), X'_n, X''_n, Y_n) \xrightarrow{w}_w ((Z|X'), (Z|(X', X'')), X', X'', Y). \tag{A.9}$$

Remark A.1 A special case with $X''_n = X'' = 1$ and $Y_n = Y = 1$ is that where $(E(z_n|X_n), X'_n) \xrightarrow{w} (E(z|X'), X')$ such that $X''_n|X'_n \xrightarrow{w}_w X''|X'$ is trivial, and hence, $E(z_n|X'_n) \xrightarrow{w} E(z|X')$ if X'_n are X_n-measurable. In terms of conditional distributions, the joint convergence $((Z_n|X_n), X'_n) \xrightarrow{w} ((Z|X'), X')$ implies that $Z_n|X'_n \xrightarrow{w}_w Z|X'$ (or more strongly, $((Z_n|X'_n), (Z_n|X_n), X'_n) \xrightarrow{w} ((Z|X'), (Z|X'), X'))$ for X_n-measurable X'_n. Another special case, clarifying the importance of the uniform integrability requirement, is $z_n = X_n = X''_n$, $z = X''$ and $X'_n = X' = Y_n = Y = 1$, where Theorem A.2 reduces to the fact that $z_n \xrightarrow{w} z$ implies $Ez_n \to Ez$ for uniformly integrable rv's z_n, z.

Remark A.2 Theorem A.2 can be applied to the bootstrap p-value. Let (A.7) hold for $z_n = p^*_n$ and $Y_n = Y = 1$, and let G^* be the conditional cdf of $p^*|(X', X'')$. If $E(G^*|X')$ equals pointwise the cdf of the $U(0,1)$ distribution, then the convergence $p^*_n|X'_n \xrightarrow{w}_w p^*|X'$ implied by Theorem A.2 under the condition $X''_n|X'_n \xrightarrow{w}_w X''|X'$ becomes $p^*_n|X'_n \xrightarrow{w}_p U(0,1)$. □

We conclude the section with a result that is used for establishing the joint convergence of original and bootstrap quantities as an implication of a marginal and a conditional convergence.

Lemma A.3 *Let $(\Omega \times \Omega^*, \mathcal{F} \times \mathcal{F}^*, P \times P^b)$ be a product probability space. Let $D_n : \Omega \to \mathcal{S}_D$, $W^*_n : \Omega^* \to \mathcal{S}_W$, $X : \Omega \to \mathcal{S}_X$ and $Z^* : \Omega \times \Omega^* \to \mathcal{S}_Z$ ($n \in \mathbb{N}$) be random elements of the Polish spaces $\mathcal{S}_D$, $\mathcal{S}_W$, $\mathcal{S}_X = \mathcal{S}'_X \times \mathcal{S}''_X$ and $\mathcal{S}_Z$. Assume further that X_n are D_n-measurable random elements of $\mathcal{S}_X$ and Z^*_n are (D_n, W^*_n)-measurable random elements of $\mathcal{S}_Z$ ($n \in \mathbb{N}$). If $X_n \xrightarrow{p} X = (X', X'')$ and $Z^*_n|D_n \xrightarrow{w}_p Z^*|X'$, then $(Z^*_n, X_n) \xrightarrow{w} (Z, X)$ and $(Z^*_n, X_n)|D_n \xrightarrow{w}_p (Z, X)|X$ on $\mathcal{S}_Z \times \mathcal{S}_X$, where Z is a random element of $\mathcal{S}_Z$ such that the conditional distributions $Z^*|X'$, $Z|X'$ and $Z|X$ are equal a.s.*

The existence of a random element Z with the specified properties, possibly on an extension of the original probability space, is ensured by Lemma 5.9 of Kallenberg (1997). A well-known special case of Lemma A.3 is that where $X_n = (1, X''_n)$, $X = (1, X'')$, $X''_n \xrightarrow{p} X''$ and $Z^*_n|X''_n \xrightarrow{w}_p X''$ (such that $D_n = X''_n$). Then $(Z^*_n, X''_n) \xrightarrow{w} (Z, X'')$ with $X'' \overset{d}{=} Z \overset{d}{=} Z|X''$ reducing to the condition that X'' and Z are independent and distributed like X'' (DasGupta, 2008, p.475).

B Proofs of the main results

Proof of Theorem 3.1. The random element (τ, F) of $\mathbb{R} \times \mathscr{D}(\mathbb{R})$ is a measurable function of (τ, X) determined up to indistinguishability by the joint distribution of (τ, X). By extended Skorokhod coupling (Corollary 5.12 of Kallenberg, 1997), we can regard the data and (τ, X) as defined on a special probability space where $(\tau_n, F^*_n) \to (\tau, F)$ a.s. in $\mathbb{R} \times \mathscr{D}(\mathbb{R})$ and $F(\cdot) = P(\tau \le \cdot|X)$ still holds. We can also replace the redefined F by a sample-path continuous random cdf that it is indistinguishable from it (and maintain the notation F).

Since F is sample-path continuous and F^*_n, F are (random) cdf's, $F^*_n \xrightarrow{a.s.} F$ in $\mathscr{D}(\mathbb{R})$ implies that $\sup_{u\in\mathbb{R}} |F^*_n(u) - F(u)| \xrightarrow{a.s.} 0$. Therefore, $F^*_n(\tau_n) - F(\tau_n) \xrightarrow{a.s.} 0$. Since $\tau_n \xrightarrow{a.s.} \tau$ and F is sample-path continuous, it holds further that $F(\tau_n) \xrightarrow{a.s.} F(\tau)$, so also $F^*_n(\tau_n) \xrightarrow{a.s.} F(\tau)$ on the special probability space. Hence, in general, $F^*_n(\tau_n) \xrightarrow{w} F(\tau)$.

Finally, we notice that $F(\tau) \sim U(0,1)$. In fact, still by the continuity of F and by the choice of F^{-1} as the right-continuous inverse, the equality of events $\{F(u) \le q\} = \{u \le F^{-1}(q)\}$, $q \in (0,1)$, holds and implies that

$$P\left(F(u)|_{u=\tau} \le q \middle| X\right) = P\left(\tau \le F^{-1}(q) \middle| X\right) = F\left(F^{-1}(q)\right)$$

as asserted, the latter equality because $F^{-1}(q)$ is X-measurable. □

Details of Remark 3.5. If $\tau^*_n \xrightarrow{w^*}_w \tau^*|X$ and $(\tau^*_n, \tau_n, X_n) \xrightarrow{w} (\tau^*, \tau, X)$ with D_n-measurable X_n ($n \in \mathbb{N}$), then the joint convergence $((\tau^*_n|D_n), \tau_n, X_n) \xrightarrow{w}_w ((\tau^*|X), \tau, X)$ follows by Lemma A.1(b). If the conditional distributions $\tau^*|X$ and $\tau|X$ are equal a.s. and F is sample-path continuous, then $(\tau_n, F^*_n) \xrightarrow{w} (\tau, F)$ by Lemma A.2(b).

Proof of Theorem 3.2. The result (as well as Corollary 3.1) follows from Theorem 3.3, which is proved below in an independent manner. Specifically, as the conditions of Theorem 3.3 are satisfied, it holds that $P(p^*_n \le q|X_n) \xrightarrow{w} F(F^{*-1}(q))$. Let $g(\cdot) = \min\{\cdot, 1\}\mathbb{I}_{\{\cdot \ge 0\}}$. By the definition of weak convergence,

$$\begin{aligned} P\left(p^*_n \le q\right) &= E\{g(P(p^*_n \le q|X_n))\} \xrightarrow{w} E\{g(F(F^{*-1}(q)))\} \\ &= E\{F(F^{*-1}(q))\} = E\{E[F(F^{*-1}(q))|F^*]\} = E\{F^*(F^{*-1}(q))\} = q \end{aligned}$$

using for the penultimate equality the F^*-measurability of $F^{*-1}(q)$ and the relation $E(F(\gamma)|F^*) = F^*(\gamma)$ for F^*-measurable rv's γ. Thus, $P(p^*_n \le q) \to q$ for almost all $q \in (0,1)$, which proves that $p^*_n \xrightarrow{w} U(0,1)$. □

Details of Remark 3.6. We justify an assertion of Remark 3.6 regarding condition (†). Let $\tilde{\tau}_n^*$ be a measurable transformation of X_n and W_n^* such that the expansion $\tau_n^* = \tilde{\tau}_n^* + o_p(1)$ holds w.r.t. the probability measure on the space where D_n and W_n^* are jointly defined. Then it holds that $(\tau_n^* - \tilde{\tau}_n^*)|D_n \xrightarrow{w}_p 0$ because convergence in probability to zero is preserved upon conditioning. As the conditional distributions $\tilde{\tau}_n^*|D_n$ and $\tilde{\tau}_n^*|X_n$ are equal a.s., it follows that the Lévy distance between $F_n^*(\cdot) := P(\tau_n^* \leq \cdot | D_n)$ and $X_n' := P(\tilde{\tau}_n^* \leq \cdot | X_n)$ is $o_p(1)$, and since the weak limit F^* of F_n^* is sample-path continuous, $F_n^* = X_n' + o_p(1)$ in the uniform distance. Thus, also $X_n' \xrightarrow{w} F^*$, such that condition (†) is satisfied with $X_n' = P(\tilde{\tau}_n^* \leq \cdot | X_n) \in \mathscr{D}(\mathbb{R})$ and $X = F^*$. □

Proof of Corollary 3.1. Convergence (3.4) implies condition (3.3) with the specified F, F^* by Lemma A.2, since the limit random measures are diffuse. Part (a) follows from Theorem 3.2 with $F = F^*$ (see Remark 3.6), and part (b) from Theorem 3.2 with $F^*(u) = E(F(u)|X')$, $u \in \mathbb{R}$. □

Proof of Theorem 3.3. Introduce $F_n(\cdot) := P(\tau_n \leq \cdot | X_n)$, $F_n^*(\cdot) := P(\tau_n^* \leq \cdot | D_n)$ and $\tilde{F}_n^*(\cdot) := P(\tau_n^* \leq \cdot | X_n)$ as random elements of $\mathscr{D}(\mathbb{R})$. On the probability space of X' (where X' is as in Theorem 3.2), possibly upon extending it, define $\tau^* := F^{*-1}(\zeta)$ for a rv $\zeta \sim U(0,1)$ which is independent of X'. Then the convergence $(F_n^*, X_n') \xrightarrow{w} (F^*, X')$, where F^* is X'-measurable and sample-path continuous, implies that $((\tau_n^*|D_n), X_n') \xrightarrow{w}_w ((\tau^*|X'), X')$ by Lemma A.2(b). Since X_n' is D_n-measurable, by Theorem A.2 (see also Remark A.1) it follows that $(\tau_n^*|D_n,\ \tau_n^*|X_n') \xrightarrow{w}_w (\tau^*|X',\ \tau^*|X')$. Since the conditional cdf F^* of $\tau^*|X'$ is sample-path continuous, for $r_n(\cdot) := F_n^*(\cdot) - P(\tau_n^* \leq \cdot | X_n')$ it follows that $\sup_{x\in\mathbb{R}} |r_n(x)| \xrightarrow{p} 0$, by using Lemma A.2(b). Then the D_n-measurability of X_n, the X_n-measurability of X_n' and Jensen's inequality yield

$$\left|\tilde{F}_n^*(u) - P(\tau_n^* \leq u|X_n')\right| = |E\{r_n(u)|X_n\}| \leq E\{|r_n(u)||X_n\} \leq E\{\sup_{x\in\mathbb{R}} |r_n(x)||X_n\}$$

for every $u \in \mathbb{R}$, and further,

$$\sup_{\mathbb{R}} \left|\tilde{F}_n^* - P(\tau_n^* \leq \cdot|X_n')\right| \leq E\{\sup_{x\in\mathbb{R}} |r_n(x)||X_n\} \xrightarrow{p} 0$$

because the $o_p(1)$ property of $\sup_{x\in\mathbb{R}} |r_n(x)|$ is preserved upon conditioning and because $\sup_{x\in\mathbb{R}} |r_n(x)|$ is bounded. Therefore, $F_n^* = P(\tau_n^* \leq \cdot|X_n') + r_n = \tilde{F}_n^* + o_p(1)$ uniformly. Then the convergence $(F_n, F_n^*) \xrightarrow{w} (F, F^*)$ in $\mathscr{D}(\mathbb{R})^{\times 2}$ extends to $(F_n, F_n^*, \tilde{F}_n^*) \xrightarrow{w} (F, F^*, F^*)$ in $\mathscr{D}(\mathbb{R})^{\times 3}$.

Fix a $q \in (0,1)$ at which F^{*-1} is a.s. continuous; such q are all but countably many because F^{*-1} is càdlàg. Here F^{*-1} stands for the right-continuous generalized inverse of F^*, and similarly for other cdf's. It follows from the CMT that $(F_n, F_n^{*-1}(q), \tilde{F}_n^{*-1}(q)) \xrightarrow{w} (F, F^{*-1}(q), F^{*-1}(q))$ in $\mathscr{D}(\mathbb{R}) \times \mathbb{R}^2$. Hence, $F_n^{*-1}(q) = \tilde{F}_n^{*-1}(q) + o_p(1)$ such that $P(|F_n^{*-1}(q) - \tilde{F}_n^{*-1}(q)| < \epsilon) \to 1$ for all $\epsilon > 0$. With $\mathbb{I}_{n,\epsilon} := \mathbb{I}_{\{|F_n^{*-1}(q) - \tilde{F}_n^{*-1}(q)| < \epsilon\}} = 1 + o_p(1)$, it holds that

$$\begin{aligned} &|P(\tau_n \leq F_n^{*-1}(q)|X_n) - P(\tau_n \leq \tilde{F}_n^{*-1}(q) + \epsilon|X_n)| \\ &\qquad \leq \mathbb{I}_{n,\epsilon}|P(\tau_n \leq \tilde{F}_n^{*-1}(q) + \epsilon|X_n) - P(\tau_n \leq \tilde{F}_n^{*-1}(q) - \epsilon|X_n)| + (1 - \mathbb{I}_{n,\epsilon}) \end{aligned}$$

$$= \mathbb{I}_{n,\epsilon}|F_n(\tilde{F}_n^{*-1}(q)+\epsilon) - F_n(\tilde{F}_n^{*-1}(q)-\epsilon)| + (1-\mathbb{I}_{n,\epsilon}),$$

the equality because $\tilde{F}_n^{*-1}(q) \pm \epsilon$ are X_n-measurable. Using the continuity of F and the CMT, we conclude that the upper bound in the previous display converges weakly to $|F(F^{*-1}(q)+\epsilon) - F(F^{*-1}(q)-\epsilon)|$, which in its turn converges in probability to zero as $\epsilon \to 0^+$ again by the continuity of F. Therefore,

$$\lim_{\epsilon\to 0^+} \limsup_{n\to\infty} P\left(\left|P(\tau_n \leq F_n^{*-1}(q)|X_n) - P(\tau_n \leq \tilde{F}_n^{*-1}(q)+\epsilon|X_n)\right| > \eta\right) = 0$$

for every $\eta > 0$. On the other hand, as it was already used, by the X_n measurability of $\tilde{F}_n^{*-1}(q)+\epsilon$ and the CMT,

$$P(\tau_n \leq \tilde{F}_n^{*-1}(q)+\epsilon|X_n) = F_n(\tilde{F}_n^{*-1}(q)+\epsilon) \xrightarrow[n\to\infty]{w} F(F^{*-1}(q)+\epsilon) \xrightarrow[\epsilon\to 0^+]{w} F(F^{*-1}(q)).$$

Theorem 4.2 of Billingsley (1968) thus yields $P(\tau_n \leq F_n^{*-1}(q)|X_n) \xrightarrow{w} F(F^{*-1}(q))$. The proof of (3.5) is concluded by noting that $P(p_n^* \leq q|X_n)$ differs from $P(\tau_n \leq F_n^{*-1}(q)|X_n)$ by no more than the largest jump of F_n^*, which tends in probability to zero because the weak limit of F_n^* is continuous.

Asymptotic validity of the bootstrap conditional on X_n requires that $F(F^{*-1}(q)) = q$ for almost all $q \in (0,1)$, which by the continuity of F and F^* reduces to $F = F^*$. □

Proof of Corollary 3.2. Part (a) follows from Theorem 3.3 with $F = F^*$ and Polya's theorem. Regarding part (b), $((\tau_n|X_n), (\tau_n^*|D_n), X_n', X_n'') \xrightarrow{w}_w (\tau|(X', X''), (\tau|X'), X', X'')$ and $X_n''|X_n' \xrightarrow{w}_w X''|X'$ imply, by Theorem A.2 with $Y_n = E\{g(\tau_n^*)|D_n\}$, $Y = E\{g(\tau)|X'\}$ and an arbitrary $g \in \mathcal{C}_b(\mathbb{R})$, that $(\tau_n|X_n', \tau_n^*|D_n) \xrightarrow{w}_w (\tau|X', \tau|X')$. As the conditional distribution $\tau|X'$ is diffuse, the proof is completed as in part (a). □

Details of Remark 3.10. With (X_n', X_n'') and (X', X'') as in Remark 3.10, and with the notation of Section S.3, we argue next that the weak convergence of $\tau_n|X_n$, $\tau_n^*|D_n$ and (X_n', X_n'') is joint. Consider a Skorokhod representation of D_n and (M, ξ_1, ξ_2) on a probability space where convergence (S.8) is strengthened to $(\tau_n, X_n', (X_n')^{1/2}X_n'')|X_n \xrightarrow{w}_{a.s.} (\tau, M, (1-\omega_{\varepsilon|\eta})M^{1/2}\xi_2)|(M, \xi_2)$ (by Lemma A.2(a)), and $\hat{\omega}_\varepsilon \xrightarrow{a.s.} \omega_\varepsilon$. Thus, on this space, $\tau_n|X_n \xrightarrow{w}_{a.s.} \tau|(M, \xi_2)$, $(X_n', X_n'') \xrightarrow{a.s.} (M, (1-\omega_{\varepsilon|\eta})\xi_2)$, and by (2.4), also $P^*(\tau_n^* \leq u) \xrightarrow{a.s.} \Phi(\omega_\varepsilon^{-1/2}M^{1/2}u)$, $u \in \mathbb{R}$, such that $\tau_n^*|D_n \xrightarrow{w}_{a.s.} \tau|M$. It follows that on a general probability space $((\tau_n|X_n), (\tau_n^*|D_n), X_n', X_n'') \xrightarrow{w}_w (\tau|(M, \xi_2), (\tau|M), M, (1-\omega_{\varepsilon|\eta})\xi_2)$. □