BOOTSTRAP INFERENCE UNDER RANDOM DISTRIBUTIONAL LIMITS

Giuseppe Cavaliere*† and Iliyan Georgiev†

Abstract

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. Nevertheless, apart from an unconditional limit distribution, a statistic of interest may possess a host of (random) conditional limit distributions. This allows the understanding of bootstrap validity to be widened, while maintaining the requirement of asymptotic control over the frequency of corrrect inferences. First, we provide conditions for the bootstrap to be asymptotically valid as a tool for conditional inference, in cases where a bootstrap distribution estimates consistently, in a sense weaker than the standard weak convergence in probability, a conditional limit distribution of a statistic. Second, we prove asymptotic bootstrap validity in a more basic, on-average sense, in cases where the unconditional limit distribution of a statistic can be obtained by averaging a (random) limiting bootstrap distribution. As an application, we establish rigorously the validity of fixed-regressor bootstrap tests of parameter constancy in linear regression models.

Keywords: Bootstrap; Random Probability Measures; Parameter Constancy Tests.
JEL Classification: C32.

1 Introduction

Asymptotic bootstrap validity is usually understood and established as consistency of the distribution of a bootstrap statistic, conditional on the data, for the unconditional limit distribution of a statistic of interest. In many applications, however, the bootstrap statistic may possess, conditionally on the data, a *random* limit distribution. Among others, cases of random bootstrap limit distributions are documented for infinite variance processes (Athreya, 1987; Knight, 1989; Aue *et al.*, 2008; Cavaliere *et al.*, 2016), time series with unit roots (Basawa *et al.*, 1991; Cavaliere *et al.*, 2015), and parameters on the boundary of the parameter space (Andrews, 2000). In most cases, the occurrence of a random limit distribution for a bootstrap statistic given the data – in contrast to a necessarily non-random limit

*Corresponding author. †Department of Statistical Sciences, University of Bologna, Italy. Correspondence to: Giuseppe Cavaliere, Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40125 Bologna, Italy; email: giuseppe.cavaliere@unibo.it.

of the unconditional distribution of the corresponding statistic, computed on the original sample – is taken as evidence of failure of the bootstrap.

In this paper we show that randomness in the limiting distribution of a bootstrap statistic need not invalidate bootstrap inference, as the bootstrap may still deliver confidence intervals (or hypothesis tests) with the desired coverage probability (or size) when the sample size diverges. Moreover, in such cases the bootstrap may also have the appealing asymptotic interpretation of a conditional inferential procedure, and may deliver efficiency (or power) gains over unconditional inference.

To see why, it is useful to note that, apart from an unconditional limit distribution, a statistic in general possesses a family of (random) conditional limit distributions, depending on the choice of the conditioning σ-algeba. If one of these conditional (random) limit distributions matches the (random) limit distributions of a chosen bootstrap statistic, then – under regularity conditions that will be discussed in the paper – inference based on the bootstrap is asymptotically valid and, importantly, conditional in nature. This observation was initially made by Lepage and Podgorski (1996) but has not been pursued further in the bootstrap literature, in particular because its development requires probabilistic tools that are not widely popular in this field.

Conditional inference could be justified by the 'conditionality principle', according to which "the evidential meaning of any outcome of any mixture experiment is the same as that of the corresponding outcome of the corresponding component experiment, ignoring the overall structure of the mixture experiment" (Birnbaum, 1962, p.271). Whenever for a statistic of interest the bootstrap estimates consistently a component of the limit unconditional distribution viewed as a mixture of conditional distributions, the bootstrap can be regarded as a large-sample implementation of the conditionality principle.

In such cases the bootstrap replicates asymptotically the property of conditional tests and confidence intervals to have conditionally constant size and coverage probability, respectively. Regarding tests, this property has been argued to be necessary for test optimality in the special case of conditioning on a complete sufficient statistic; see Lockhart (2012). More generally, gains in power and precision can be expected to occur when the reference population is effectively restricted to outcomes that share statistically relevant features with the actual sample. For instance, in the case of confidence intervals, Lepage and Podgorski (1996, Figure 2) provide numerical evidence of substantial precision gains in a particular implementation of a permutation bootstrap with a random limit of the bootstrap statistic, where conditioning is on the order statistics of regression residuals (ancillary in that context).

Following a practice in the literature (see, e.g., Lockhart, 2012), we recast the constant conditional size and coverage probability property into the requirement that bootstrap p-values should be uniformly distributed conditionally, at least asymptotically. One of our main results is a general sufficient condition for this to be the case. We also provide conditions for the more basic property of unconditional asymptotic distributional uniformity of p-values; this property implies asymptotic control of the frequency of wrong inferences on average over the conditioning variables but no longer warrants a conditional interpretation of the bootstrap inferential procedure.

When dealing with random limiting distributions, the usual convergence concept employed to establish bootstrap validity, i.e. weak convergence in probability, can only be employed in some special cases. Therefore, in this paper we discuss asymptotic bootstrap validity also in cases where consistency of the bootstrap distribution for a conditional (null) limit distribution of an original statistic holds in a sense weaker than weak convergence in probability.

To show the practical relevance of our results, we include an analysis of the well-known and much applied (but also misunderstood) tests of parameter constancy in regression models where the design matrix could be random but be conditioned upon; see Hall (1991,p.170). In the resampling process forming the bootstrap sample, it appears natural to take the design matrix as fixed, i.e. it does not vary across the bootstrap repetitions. Accordingly, bootstrap algorithms with this feature are sometimes labelled as 'fixed design', 'fixed regressor' or 'conditional' bootstrap. Under a set of assumptions proposed by Hansen (2000), we argue that the fixed-regressor bootstrap test statistics have random limit distributions, thus invalidating previous claims that the bootstrap is consistent for the unconditional limit distribution of the original parameter constancy test statistics. Then we provide conditions under which the fixed-regressor bootstrap entails conditional asymptotic inference.

The paper is organized as follows. Before presenting the theoretical results, in Section 2 we discuss a sequence of simplified preliminary examples, including – among others – (Kolmogorov-Smirnov type) tests of correct distributional specification and CUSUM tests under infinite variance. Our main theoretical results are presented in Section 3. In Section 4 we discuss bootstrap tests of parameter constancy. Auxiliary proofs are reported in Section 6, whereas proofs of the claims in Section 4 are collected in an on-line supplement.

Notation and preliminary assumptions

We use the following notation throughout. The Skorokhod spaces of càdlàg functions $\mathbb{R} \to \mathbb{R}$, $[0,1] \to \mathbb{R}^{m\times n}$ and $[0,1] \to \mathbb{R}^n$ are denoted by $\mathscr{D}(\mathbb{R})$, $\mathscr{D}_{m\times n}$ and $\mathscr{D}_n$, respectively; for the latter, when $n = 1$ the subscript 1 is suppressed. Integrals are over $[0,1]$ unless otherwise stated, Φ is the standard Gaussian cumulative distribution function [cdf], $U(0,1)$ is the uniform distribution on $[0,1]$ and $\mathbb{I}_{(.)}$ is the indicator function.

We assume always that well-defined conditional distributions exist. Whenever interest is in the 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, Th. 5.3). For random cdf's on $\mathbb{R}$ (resp., for the underlying conditional distributions) equalities are understood up to indistinguishability.

For random elements (Z,Y), (Z_n,Y_n) of some metric spaces $\mathcal{S}'\times\mathcal{S}''$ and $\mathcal{S}'\times\mathcal{S}''_n$ $(n\in\mathbb{N})$, and defined on a common probability space, we denote by $Z_n|Y_n \xrightarrow{w}_p Z|Y$ (resp. $Z_n|Y_n \xrightarrow{w}_{a.s.} Z|Y$) the fact that $E\left(g\left(Z_n\right)|Y_n\right) \to E\left(g\left(Z\right)|Y\right)$ in probability (resp. a.s.) for all continuous bounded functions $g : \mathcal{S}' \to \mathbb{R}$. Whenever Z_n and Z are scalar random variables [rv's], this is equivalent to the convergence $P\left(Z_n \le \cdot|Y_n\right) \to P\left(Z \le \cdot|Y\right)$ of the random cdf's, in probability (resp. a.s.) in $\mathscr{D}(\mathbb{R})$. Further, whenever $P\left(Z \le \cdot|Y\right) = P\left(Z \le \cdot\right)$, it reduces to the concept of weak convergence in probability (resp. a.s.) usually employed in the bootstrap literature. By extension, we use the same terminology also when the random process $P\left(Z \le \cdot|Y\right)$ has a non-degenerate distribution.

On the other hand, if (Z,Y), (Z_n,Y_n) $(n\in\mathbb{N})$ are defined on possibly different probability spaces, we denote by $Z_n|Y_n \xrightarrow{w}_w Z|Y$ the fact that $E(g(Z_n)|Y_n) \xrightarrow{w} E(g\left(Z\right)|Y)$ for all continuous bounded functions $g : \mathcal{S}' \to \mathbb{R}$, which corresponds to the probabilistic concept of weak convergence of random measures (here, of the random conditional distributions $Z_n|Y_n$; see Daley and Vere-Jones, 2008, p.138). If on another probability space we are given $(\tilde{Z}, \tilde{Y})$ distributed like (Z, Y) and if $Z_n|Y_n \overset{w}{\to}_w Z|Y$, then $Z_n|Y_n \overset{w}{\to}_w \tilde{Z}|\tilde{Y}$ because $E(g(\tilde{Z})|\tilde{Y})$ has the same distribution as $E(g(Z)|Y)$.

Suppose now that $Z_n = (Z'_n, Z''_n)$, $Y_n = (Y'_n, Y''_n)$, and similarly for Z and Y. We say that $Z'_n|Y'_n \overset{w}{\to}_w Z'|Y'$ and $Z''_n|Y''_n \overset{w}{\to}_w Z''|Y''$ *jointly* (denoted also by $(Z'_n|Y'_n, Z''_n|Y''_n) \overset{w}{\to}_w (Z'|Y', Z''|Y''))$ if

$$(E(h(Z'_n)|Y'_n), E(k(Z''_n)|Y''_n)) \overset{w}{\to} (E(h(Z')|Y'), E(k(Z'')|Y'')) \tag{1}$$

for all continuous and bounded real functions h and k with matching domain. Even for $Y'_n = Y''_n$, this is distinct from the convergence $(Z'_n, Z''_n)|Y'_n \overset{w}{\to}_w (Z', Z'')|Y$ defined by $E(g(Z'_n, Z''_n)|Y'_n) \overset{w}{\to} E(g(Z', Z'')|Y)$ for all continuous and bounded $g : \mathcal{S}' \times \mathbb{R}$. If Z'_n, Z''_n, Z' and Z'' are rv's, (1) is equivalent to the weak convergence of the associated random cdf's as random elements of $\mathscr{D}(\mathbb{R}) \times \mathscr{D}(\mathbb{R})$:

$$(P(Z'_n \leq \cdot|Y'_n), P(Z''_n \leq \cdot|Y''_n)) \overset{w}{\to} (P(Z' \leq \cdot|Y'), P(Z'' \leq \cdot|Y''));$$

cf. Daley and Vere-Jones (2008, pp.143-144).

On probability spaces where both the statistical data and the auxiliary variates used in the construction of the bootstrap data are defined, we use $Z_n \overset{w^*}{\to}_p Z|Y$ (resp. $\overset{w^*}{\to}_{a.s}$, $\overset{w^*}{\to}_w$) interchangeably with $Z_n|Y_n \overset{w}{\to}_p Z|Y$ (resp. $\overset{w}{\to}_{a.s}$, $\overset{w}{\to}_w$), and write $P^*(\cdot)$ for $P(\cdot|Y_n)$, provided that $\sigma(Y_n)$ coincides with the σ-algebra induced by the original data.

2 Examples

We anticipate our main results in a sequence of stylized examples. These serve as a vehicle to make three important points. One point is that bootstrap validity in a sense as strong as exact conditional inference can be accompanied by consistency properties of the bootstrap formulated in terms of rather weak convergence concepts and involving random limits (Sections 2.1, 2.2). A further point is that reasoning can be inverted and such consistency properties can be used in order to establish the asymptotic validity of the bootstrap as a tool for conditional inference (Section 2.3). A final point is that even when a conditional interpretation of bootstrap inference is not warranted, the bootstrap may still be asymptoticaly valid in a more basic, on-average, sense (Sections 2.3, 2.4).

2.1 A Gaussian regression with (possibly) non-ergodic regressors

Consider a simple linear model

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

where $\{\varepsilon_t\}$ are i.i.d. N$(0, 1)$ and $\{x_t\}$ are observable random variables, independent of the unobservable ε_t's. We assume further that $M_n := \sum_{t=1}^n x_t^2 > 0$ a.s. for all n. Interest is in inference on β; for instance, a test of a null hypothesis of the form $\mathsf{H}_0 : \beta = 0$. Let $\hat{\beta}$ be the OLS estimator of β. It is immediate to see that, given the x_t's, $\hat{\beta} - \beta = M_n^{-1} \sum_{t=1}^n x_t \varepsilon_t$ is Gaussian with mean 0 and variance M_n^{-1}, i.e. $P(\hat{\beta} - \beta \le u | X_n) = \Phi(M_n^{1/2} u)$, $u \in \mathbb{R}$ (X_n being a shortcut for $\{x_t\}_{t=1}^n$). For illustrational purposes, we do not consider the studentized case.

The classical (non-parametric) fixed-design bootstrap sample is[1]

$$y_t^* = \hat{\beta} x_t + \varepsilon_t^* \qquad (t = 1, 2, ..., n)$$

where ε_t^* is i.i.d. $\mathrm{N}(0,1)$, independent of the original data. Then, if $\hat{\beta}^*$ is the OLS estimator of β from the bootstrap sample, it holds that, conditionally on the original data, $\hat{\beta}^* - \hat{\beta} \sim N(0, M_n^{-1})$ and $P^*(\hat{\beta}^* - \hat{\beta} \le u) = \Phi(M_n^{1/2} u)$, $u \in \mathbb{R}$, where P^* denotes probability conditional on the data. That is, the bootstrap statistic $\hat{\beta}^* - \hat{\beta}$ has the same distribution conditional on the data (and, in fact, on X_n alone), as the original $\hat{\beta} - \beta$ conditional on X_n: $P^*(\hat{\beta}^* - \hat{\beta} \le u) = P(\hat{\beta} - \beta \le u | X_n)$ for all real u. A few comments follow.

(i) In finite samples, bootstrap inference on β (i.e., inference using the distribution of $\hat{\beta}^* - \hat{\beta}$ under P^* as reference for $\hat{\beta} - \beta$) is *exact* and this is due to the equality of two *conditional* distributions.

(ii) Let $n^{-\alpha} M_n \xrightarrow{p} M$ as $n \to \infty$, for some positive real constants α and M. Define $\tau_n := n^{\alpha/2}(\hat{\beta} - \beta)$ and $\tau_n^* := n^{\alpha/2}(\hat{\beta}^* - \hat{\beta})$. Then the equality $P(\tau_n \le u | X_n) = P^*(\tau_n^* \le u) = \Phi(n^{-\alpha/2} M_n^{1/2} u)$ for all $u \in \mathbb{R}$ restates the exactitude of bootstrap inference, which is invariant to non-random scaling. When $n \to \infty$, it holds that $\Phi(n^{-\alpha/2} M_n^{1/2} u) \xrightarrow{p} \Phi(M^{1/2} u)$, which implies that $\tau_n | X_n \xrightarrow{w}_p N(0, M^{-1})$ and $\tau_n^* \xrightarrow{w^*}_p N(0, M^{-1})$. Being $N(0, M^{-1})$ a non-random distribution, the conditioning of τ_n on X_n is asymptotically negligible, in the sense that also the unconditional convergence $\tau_n \xrightarrow{w} N(0, M^{-1})$ holds. Therefore, the bootstrap (conditional) distribution of τ_n^* estimates consistently the limiting (unconditional) distribution of τ_n, and by continuity of Φ, the bootstrap possesses the usual validity property $\sup_{u \in \mathbb{R}} |P^*(\tau_n^* \le u) - P(\tau_n \le u)| \xrightarrow{p} 0$.

(iii) Let $n^{-\alpha} M_n \xrightarrow{p} M > 0$ a.s. as $n \to \infty$, where M is now stochastic and non-degenerate. Then, although bootstrap inference is still exact, the bootstrap no longer estimates consistently the limit distribution of τ_n, and hence, fails to be valid in the usual sense. To see this, notice that $P(\tau_n \le u)$ converges to the *non-stochastic* cdf of a mixed Gaussian distribution with mixing conditional variance M^{-1}, whereas $\sup_{u \in \mathbb{R}} |P^*(\tau_n^* \le u) - \Phi(M^{1/2} u)| \xrightarrow{p} 0$, which implies that $P^*(\tau_n^* \le \cdot)$ converges (uniformly in probability) to the *random* cdf $\Phi(M^{1/2}(\cdot))$. Hence, $\sup_{u \in \mathbb{R}} |P^*(\tau_n^* \le u) - P(\tau_n \le u)|$ is not $o_p(1)$. However, since it also holds that $\sup_{u \in \mathbb{R}} |P(\tau_n \le u | X_n) - \Phi(M^{1/2} u)| \xrightarrow{p} 0$, the bootstrap is consistent for a limiting *conditional* distribution of τ_n. This result can be formalized as an 'in probability' convergence of random measures on $\mathbb{R}$: $\tau_n | X_n \xrightarrow{w}_p N(0, M^{-1}) | M$ and $\tau_n^* | D_n \xrightarrow{w}_p N(0, M^{-1}) | M$, where $D_n := \{x_t, y_t\}_{t=1}^n$ denotes the data.

(iv) Finally, if $n^{-\alpha} M_n \xrightarrow{w} M > 0$ a.s. as $n \to \infty$, where M is stochastic and non-degenerate, then $\tau_n | X_n$ and $\tau_n^* | D_n$ need not converge weakly in probability, even to random distributions.

[1]The value of 0 could be used instead of $\hat{\beta}$, in which case $\hat{\beta}^*$ would replace $\hat{\beta}^* - \hat{\beta}$ in what follows.

Nevertheless, the exactness of bootstrap inference for finite n still implies a consistency property of the bootstrap which can be formulated as

$$\left(P\left(\tau_n \leq \cdot|X_n\right),\, P^*\left(\tau_n^* \leq \cdot\right)\right)' \xrightarrow{w} (1,1)'\Phi(M^{1/2}\cdot) \tag{3}$$

in the sense of weak convergence of random elements of $\mathscr{D}(\mathbb{R}) \times \mathscr{D}(\mathbb{R})$, or equivalently

$$(\tau_n|X_n,\, \tau_n^*|D_n)' \xrightarrow{w}_w (1,1)'N(0, M^{-1})|M \tag{4}$$

in the sense of a joint weak convergence of random measures on $\mathbb{R}$. Thus, the bootstrap is again consistent for a limiting *conditional* distribution of τ_n but the bootstrap and the original statistic converge to this joint limit in a weaker sense than in case (iii).

In summary, the fact that the bootstrap distribution is random in the limit, and hence, the bootstrap is not consistent (weakly in probability) for the unconditional limit distribution of a test statistic, does *not* imply that the bootstrap does not provide valid (in this example, even exact) inference. It will be shown later in the paper that in more general settings, where bootstrap inference is not exact, it can still be asymptotically valid as a *consequence* of consistency properties as weak as (3) or (4).

2.2 A permutation CUSUM test under infinite variance

Neither the fixed-design construction of the bootstrap statistic nor gaussianity are essential for the comments to the previous example. In fact, consider a standard CUSUM test for the null hypothesis (say, H_0) that $\{\varepsilon_t\}_{t=1}^n$ are i.i.d. random variables. The test statistic is of the form

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

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 $\nu_n = \max_{t=1,\dots,n} |\varepsilon_t|$ when $E\varepsilon_t^2 = \infty$. In particular, when ε_t is in the domain of attraction of a strictly α-stable law with $\alpha \in (0,2)$, the asymptotic distribution of τ_n depends on unknown parameters (e.g. the characteristic exponent α), which makes the test difficult to apply. To overcome this problem, Aue *et al.* (2008) consider a permutation-bootstrap analogue of τ_n, defined as

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

where π is a (uniformly distributed) random permutation of $\{1, 2, ..., n\}$, independent of the data.[2] The results in Aue *et al.* (2008) imply that, under H_0 and if ε_t is in the domain of attraction of a strictly α-stable law, $\alpha \in (0,2)$, it holds that $\tau_n \xrightarrow{w} \rho_\alpha(S)$ and $\tau_n^* \xrightarrow{w^*}_w \rho_\alpha(S)|S$ for a certain random function ρ_α 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 ρ_α independent of S.

[2] 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 $\nu_n = 1$.

(i) As in the previous example, also here the limit distribution of the bootstrap statistic τ_n^* is random. As noticed in Aue *et al.* (2008, pp.128-129), this implies that the bootstrap does not provide consistent estimation of the unconditional limit distribution of τ_n.

(ii) Aue *et al.* (2008) do not notice that the failure of the bootstrap to estimate consistently the distribution of $\rho_\alpha(S)$ does not invalidate bootstrap inference. In fact, the situation is similar to case (iv) of Section 2.1. Let X_n be the vector of order statistics of $\{\varepsilon_t\}_{t=1}^n$; then, under H_0, $\tau_n|X_n \overset{d}{=} \tau_n^*|\{\varepsilon_t\}_{t=1}^n$. As a consequence, under H_0 the permutation bootstrap implements *exact* finite-sample inference conditional on X_n. Furthermore, as $n \to \infty$, the bootstrap estimates consistently the limit of the conditional distribution $\tau_n|X_n$, in the sense of joint weak convergence of random measures (see eq. (1) for the definition):

$$(\tau_n|X_n, \tau_n^*|\{\varepsilon_t\}_{t=1}^n)' \overset{w}{\to}_w (1,1)'\rho_\alpha(S)|S \ . \tag{5}$$

(iii) CUSUM tests can be applied to residuals in order to test for correct model specification or stability of model parameters (see e.g. Ploberger and Krämer, 1992). Consider thus the case where ε_t is the unobservable disturbance component of a statistical model (as in the regression model of Section 2.1) and we only have available residuals $\hat{\varepsilon}_t$ obtained upon estimation of the model using a sample D_n, say. The CUSUM statistic is $\hat{\tau}_n := \hat{\nu}_n^{-1} \max_{t=1,\dots,n} |\sum_{i=1}^t (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}_n)|$, 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, and the bootstrap statistic is $\hat{\tau}_n^* := \hat{\nu}_n^{-1} \max_{t=1,\dots,n} |\sum_{i=1}^t (\hat{\varepsilon}_{\pi(i)} - \bar{\hat{\varepsilon}}_n)|$. It could be shown that if $\hat{\tau}_n - \tau_n \overset{p}{\to} 0$ and $(\hat{\tau}_n^* - \tau_n^*)|D_n \overset{w}{\to}_p 0$ under H_0 (e.g., due to consistent parameter estimation), then the bootstrap is consistent in the sense that

$$(\hat{\tau}_n|X_n, \hat{\tau}_n^*|D_n) \overset{w}{\to}_w (1,1)'\rho_\alpha(S)|S. \tag{6}$$

As a consequence, under the conjecture that $P(\rho_\alpha(S) \leq \cdot|S)$ defines a continuous stochastic process, the bootstrap p-value associated to the residual-based CUSUM statistic, $p_n^* := P(\hat{\tau}_n^* \leq \hat{\tau}_n|\, D_n)$, is asymptotically $U(0,1)$ distributed, both conditionally on X_n and on average, by Corollary 1 in Section 3.1. The next example sheds more light on the implication from $\overset{w}{\to}_w$ bootstrap consistency to the asymptotic uniformity of bootstrap p-values in a situation where bootstrap p-values are analysed without recourse to Corollary 1.

2.3 A cointegrating regression

A well-known example of a relation like (2) in the context of (non-stationary) time-series data is the classic co-integrating regression where $\{x_t\}$ is a unit-root process and ε_t need no longer be Gaussian with known moments and independent of $\{x_t\}$. In this case, the bootstrap introduced in Section 2.1 still features a random limit distribution but no longer delivers exact finite-sample inference.

Specifically, let $x_t = \sum_{s=1}^{t-1} \eta_s$ be the non-stationary regressor in (2) and $e_t := (\varepsilon_t, \eta_t)'$ be a stationary and ergodic martingale difference sequence [mds] with p.d. variance matrix $\Omega = \text{diag}\{\omega_{\varepsilon\varepsilon}, \omega_{\eta\eta}\}$.[3] Then $n^{-1/2} \sum_{t=1}^{\lfloor n\cdot \rfloor} e_t \overset{w}{\to} (B_\varepsilon, B_\eta)'$ in $\mathscr{D}_2$, where $(B_\varepsilon, B_\eta)'$ is a bivariate

[3]Non-diagonal Ω could be handled by either augmenting the estimated regression with Δx_{t+1} (Saikkonen, 1991) or by modifying the definition of the bootstrap errors. Since for our purposes the case of non-diagonal Ω is not qualitatively different, we stick to diagonal Ω.

Brownian motion with covariance matrix Ω; see e.g. Chan and Wei (1988). Let $\hat{\beta}$ be the OLS estimator of β, with associated residual variance $\hat{\omega}_{\varepsilon\varepsilon} := n^{-1}\sum_{t=1}^{n}(y_t - \hat{\beta}x_t)^2$. It is well-known that

$$\tau_n := n(\hat{\beta} - \beta) \xrightarrow{w} \left(\int B_\eta^2\right)^{-1} \int B_\eta dB_\varepsilon \stackrel{d}{=} N(0, \omega_{\varepsilon\varepsilon}M^{-1}) \tag{7}$$

with $M := \int B_\eta^2$, the limit being mixed Gaussian due to the independence of B_η and B_ε.

Consider the following bootstrap: $y_t^* = \hat{\beta}x_t + \hat{\omega}_{\varepsilon\varepsilon}^{1/2}\varepsilon_t^*$ $(t = 1, ..., n)$ with ε_t^* i.i.d. $N(0,1)$ and independent of the data. The bootstrap analogue of $\hat{\beta}$, say $\hat{\beta}^*$, obtains by regression of y_t^* on x_t and behaves similarly to case (iv) of Section 2.1. That is, with $\tau_n^* := n(\hat{\beta}^* - \hat{\beta})$, we have that $\tau_n^*|D_n \sim N(0, n\hat{\omega}_{\varepsilon\varepsilon}M_n^{-1})|\,(M_n, \hat{\omega}_{\varepsilon\varepsilon})$ where $D_n := \{x_t, y_t\}_{t=1}^n$ and $n^{-2}M_n := n^{-2}\sum_{t=1}^n x_t^2 \to_w M$ jointly with (7). Provided $\hat{\omega}_{\varepsilon\varepsilon} \xrightarrow{p} \omega_{\varepsilon\varepsilon}$, it follows that

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

by the continuous mapping theorem [cmt]. Hence, the bootstrap distribution has a random limit:

$$\tau_n^* \xrightarrow{w_*}_w N(0, \omega_{\varepsilon\varepsilon}M^{-1})\big|\,M \tag{8}$$

as a weak convergence of random measures, the weakest convergence concept listed in Section 2.1. The following points can be made out of this example.

(i) The unconditional limit of $\tau_n := n(\hat{\beta}-\beta)$, see eq. (7), can be recovered by integrating over M the conditional limit of $\tau_n^* := n(\hat{\beta}^* - \hat{\beta})$ given the data. This property implies asymptotic validity of the bootstrap in a basic sense that we will discuss under high level assumptions in Section 3.2. To prepare the discussion, let $p_n^* := P^*(\tau_n^* \leq \tau_n)$ be the bootstrap p-value. As τ_n is D_n-measurable, it holds that

$$\begin{aligned} p_n^* &= P^*(\tau_n^* \leq u)|_{u=\tau_n} = \Phi(\hat{\omega}_{\varepsilon\varepsilon}^{-1/2}M_n^{1/2}(\hat{\beta} - \beta)) \\ &\xrightarrow{w} \Phi((\omega_{\varepsilon\varepsilon}{\textstyle\int} B_\eta^2)^{-1/2}{\textstyle\int} B_\eta dB_\varepsilon) \stackrel{d}{=} \Phi(N(0,1)) \stackrel{d}{=} U(0,1). \end{aligned}$$

Hence, when inference (e.g. hypothesis testing) on β is based on the distribution of τ_n^* conditional on the data, the frequency of wrong inferences can be controlled in large samples. However, bootstrap inference cannot be guaranteed to have a conditional interpretation asymptotically, as the convergence of p_n^* occurs on average over $X_n := \{x_t\}_{t=1}^n$, though not necessarily conditionally on X_n (point (iv) below provides further elaboration).

(ii) The stronger result that bootstrap inference is asymptotically valid conditionally on X_n can be obtained upon a strenghtening of our previous assumptions. For instance, assume that ε_t is an mds with respect to $\mathcal{G}_t = \sigma(\{\varepsilon\}_{s=-\infty}^t \cup \{\eta_s\}_{s\in\mathbb{Z}})$, and that $n^{-1}\sum_{t=1}^n E(\varepsilon_t^2|\{\eta_s\}_{s\in\mathbb{Z}}) \xrightarrow{a.s.} \omega_{\varepsilon\varepsilon}$. It then follows (by using Theorem 5 of Georgiev *et al.*, 2016) that

$$\begin{aligned} (\tau_n|X_n, \tau_n^*|D_n)' &\xrightarrow{w}_w (1,1)'\left(\int_0^1 B_\eta^2\right)^{-1}\int_0^1 B_\eta dB_\varepsilon\Big|\,B_\eta \\ &\stackrel{d}{=} (1,1)'N(0, \omega_{\varepsilon\varepsilon}M^{-1})\big|\,M \end{aligned} \tag{9}$$

in the sense of eq. (1) or, equivalently,

$$(P(\tau_n \le \cdot|X_n),\ P^*(\tau_n^* \le \cdot))\ ' \overset{w}{\to} \Phi(\omega_{\varepsilon\varepsilon}^{-1/2}M^{1/2}(\cdot))(1,1)'$$

in $\mathscr{D}(\mathbb{R})\times\mathscr{D}(\mathbb{R})$. Hence, the bootstrap is consistent for the limiting *conditional* distribution of $\tau_n|X_n$. By the continuity of Φ, this implies that

$$\sup_{u\in\mathbb{R}}|P^*(\tau_n^* \le u) - P(\tau_n \le u|X_n)| \overset{p}{\to} 0.$$

In contradistinction to Section 2.1, here the distributional proximity of τ_n given X_n and τ_n^* given the data follows from their proximity to a common random limit (and not vice versa). In Theorem 1 we will obtain some general inferential implications of such a proximity, that we are next anticipating using the conditional normality of τ_n^* as a shortcut.

(iii) In terms of bootstrap p-values, the result in (ii) implies that

$$p_n^*|X_n = \Phi(\hat{\omega}_{\varepsilon\varepsilon}^{-1/2}M_n^{1/2}(\hat{\beta}-\beta))|X_n \overset{w}{\to}_w \Phi\left(N(0,1)\right) \overset{d}{=} U(0,1)$$

because $\hat{\omega}_{\varepsilon\varepsilon}^{-1/2}M_n^{1/2}(\hat{\beta}-\beta)|X_n \overset{w}{\to}_w (\omega_{\varepsilon\varepsilon}\int B_\eta^2)^{-1/2}\int B_\eta dB_\varepsilon|B_\eta \overset{d}{=} N(0,1)$ by (9), the X_n-measurability of M_n and the continuity of Φ. As asymptotic uniformity of p-values holds conditionally, bootstrap inference will be asymptotically valid not only on average over X_n, but also conditionally on X_n. Remarkably, in Section 3 we will show that such a strong validity property results solely from a weak consistency property of the bootstrap (such as (9)) and from the continuity of the limit random distribution.

(iv) The form of the unconditional limit of τ_n, see (7), does not imply that a conditional result like eq. (9) should hold, unless the original assumptions are strengthened as in point (ii). Hence, in general the bootstrap p-value p_n^* can be asymptotically uniform without being asymptotically conditionally uniform. To see this fact, consider the case where $\eta_t = \xi_t(1 + \mathbb{I}_{\{\varepsilon_t<0\}})$, with $\{\varepsilon_t\}$ and $\{\xi_t\}$ two independent i.i.d. sequences of zero mean, unit variance rv's. Then, $e_t = (\varepsilon_t, \eta_t)'$ is a zero-mean i.i.d. sequence with covariance $\Omega = \text{diag}\{1, 2.5\}$ and, as derived in (i), (7) holds, yielding $p_n^* \overset{w}{\to} U(0,1)$. However, in applying an invariance principle to $\sum_{s=1}^t \varepsilon_t$ conditionally on X_n, it should be taken into account that η_t is informative of the sign of ε_t; hence the ε_t's, conditionally on their own past and the whole sequence $\{x_t\}_{t=1}^n$, do not form an mds. As a consequence, it is shown in Section 6 that

$$\begin{aligned}\tau_n|X_n &\overset{w}{\to}_w \left(\int B_\eta^2\right)^{-1}\int B_\eta d\{\sqrt{\omega_{\varepsilon|\eta}}B_{y1} + \sqrt{1-\omega_{\varepsilon|\eta}}B_{y2}\} \\ &\overset{d}{=} M^{-1/2}(\sqrt{\omega_{\varepsilon|\eta}}\xi_1 + \sqrt{1-\omega_{\varepsilon|\eta}}\xi_2)\big|\, M, \xi_2 \end{aligned} \quad (10)$$

where (B_{y1}, B_{y2}, B_x) is a standard trivariate Brownian motion, $\omega_{\varepsilon|\eta} := Var(\varepsilon_s|\eta_s) \in (0,1)$, and M, ξ_1, ξ_2 are jointly independent with $\xi_i \sim N(0,1), i = 1, 2$. The bootstrap, instead of estimating consistently the limiting conditional distribution of τ_n given X_n, estimates a mixture distribution obtained by integrating ξ_2 out of this limit. As a result, conditionally on X_n the bootstrap p-value is not asymptotically uniformly distributed:

$$p_n^*|X_n = \Phi(\hat{\omega}_{\varepsilon\varepsilon}^{-1/2}M_n^{1/2}(\hat{\beta}-\beta))\Big|\, X_n \overset{w}{\to}_w \Phi(\sqrt{\omega_{\varepsilon|\eta}}\xi_1 + \sqrt{1-\omega_{\varepsilon|\eta}}\xi_2)\big|\, \xi_2, \quad (11)$$

which does not correspond to a $U(0,1)$ distribution.

2.4 A parametric bootstrap goodness-of-fit test

The parametric bootstrap is among the usual techniques 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), there exist relevant settings where a random limit of $n^{-\alpha}M_n$ implies that parametrically bootstrapped goodness-of-fit test statistics have random limit distributions.

Let the null hypothesis of interest, say H_0, be that the standardized errors ε_t/σ in (2) have certain known density f with mean 0 and variance 1. For expositional ease we assume that $\sigma = 1$ and is known to the statistician. Then the Kolmogorov-Smirnov statistic based on OLS residuals $\hat{\varepsilon}_t$ is $\tau_n := n^{1/2}\sup_{u\in\mathbb{R}}|n^{-1}\sum_{t=1}^n \mathbb{I}_{\{\hat{\varepsilon}_t\le u\}} - \int_{-\infty}^u f|$. 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 residuals $\hat{\varepsilon}_t^*$; and (iii) calculating τ_n^* as $\tau_n^* := n^{1/2}\sup_{u\in\mathbb{R}}|n^{-1}\sum_{t=1}^n \mathbb{I}_{\{\hat{\varepsilon}_t^*\le u\}} - \int_{-\infty}^u f|$.

To see that the distribution of the bootstrap statistic τ_n^* conditional on the data may have a random limit, consider the Gaussian case, $f = \Phi'$, and let $\mathbb{P}$ be the measure in the product probability space on which the data and $\{\varepsilon_t^*\}$ are jointly defined. Under the assumptions of Johansen and Nielsen (2016, Sec. 4.1-4.2), it holds (*ibidem*) that $\tau_n^* = \tilde{\tau}_n^* + o_{\mathbb{P}}(1)$ with

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

where $q(u) = \Phi^{-1}(u)$ is the u-th quantile of Φ. The expansion of τ_n^* holds also conditionally on the data $D_n := \{x_t, y_t\}_{t=1}^n$, 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 in the sense of eq. (1), so does $\tau_n^*|\,D_n$ for the same limit. Assume that $n^{-\alpha/2}x_{\lfloor n\cdot\rfloor} \xrightarrow{w} X_\infty$ in $\mathscr{D}$ for some $\alpha > 0$ and that $M := \int X_\infty^2 > 0$ a.s. (e.g., $X_\infty = B_\eta$ for x_t defined as in Section 2.3); then $(M_n, \xi_n) := (\sum_{t=1}^n x_t^2, \sum_{t=1}^n x_t)$ satisfies $(n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n) \xrightarrow{w} (M, \xi := \int X_\infty)$. Since M_n and ξ_n are functions of the data, this convergence is equivalent to $(n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n) \xrightarrow{w^*}_w (M, \xi)|M, \xi$. Furthermore, if $W_n^*(u) := n^{-1}\sum_{t=1}^n(\mathbb{I}_{\{\varepsilon_t^*\le q(u)\}} - u)$, $u \in [0,1]$, 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 standard Gaussian), but not jointly independent of the data, because

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

$u \in [0,1]$, where $\psi(\cdot) := E^*[\varepsilon_1^*\mathbb{I}_{\{\varepsilon_1^*\le q(\cdot)\}}] = -\Phi'(q(\cdot))$ is a trimmed mean function. It is argued in Section 6 that, more strongly,

$$(n^{1/2}W_n^*,\, n^{(\alpha+1)/2}\hat{\beta}^*,\, n^{-\alpha/2-1}\xi_n) \xrightarrow{w^*}_w \left(W, M^{-1/2}b, \xi\right)\big|\, M, \xi \tag{12}$$

on $\mathscr{D}\times\mathbb{R}^2$, where $(W,b)|M,\xi$ is a pair of a standard Brownian bridge and a standard Gaussian rv individually independent of M, ξ but with Gaussian joint conditional distributions having covariance $\text{Cov}(W(u), b|M, \xi) = M^{-1/2}\xi\psi(u)$, $u \in [0,1]$. Combining these pieces with the cmt yields

$$\tau_n^* \xrightarrow{w^*}_w \sup_{u\in[0,1]} |W(u) + \Phi'(q(u))M^{-1/2}b\xi|\,|M, \xi \overset{d}{=} \sup_{u\in[0,1]} |\tilde{W}(u)|\,|M, \xi\ , \tag{13}$$

where, conditionally on M, ξ, $\tilde{W}$ is a zero-mean Gaussian process with $\tilde{W}(0) = \tilde{W}(1) = 0$ a.s. and conditional covariance function $K(u, v) = u(1-v) - M^{-1}\xi^2\Phi'(q(u))\Phi'(q(v))$ for $0 \le u \le v \le 1$. In summary, the limit bootstrap distribution is random because the latter conditional covariance is random whenever M or ξ are such.

We now discuss whether, an in what sense, τ_n^* can provide a distributional approximation of τ_n.

(i) As for τ_n^*, under the null that $\varepsilon_t \sim i.i.d.N(0,1)$, 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 $\tilde{\tau}_n := \sup_{u\in[0,1]} |n^{-1/2}\sum_{t=1}^n(\mathbb{I}_{\{\varepsilon_t\le q(u)\}} - u) + \Phi'(q(u))\hat{\beta}n^{-1/2}\sum_{t=1}^n x_t|$ defined similarly to $\tilde{\tau}_n^*$. Hence, it is possible for $\tau_n|X_n$ (with $X_n := \{x_t\}_{t=1}^n$) to have the same random limit distribution₋ as τ_n^* given the data (for instance, if the null hypothesis is true and $\{\varepsilon_t\}$ is an i.i.d. sequence independent of $\{x_t\}$). In such a case, as we will show in Theorem 1, bootstrap p-values are asymptotically $U(0,1)$ conditionally on X_n. However, it is implausible that $\tau_n|X_n \xrightarrow{w}_w \tau_\infty|M, \xi$ in general (for example under the assumptions that induced (10) of Section 2.3), even if the ε_t's are Gaussian. As a result, bootstrap inference cannot be expected to be in general valid conditionally on X_n.

(ii) Under H_0, however, bootstrap inference is valid at least in the basic sense of keeping controlled the large-sample frequency of incorrect inferences. It is readily seen that $\tau_n = \tilde{\tau}_n + o_p(1) \xrightarrow{w} \sup_{u\in[0,1]} |\tilde{W}(u)| =: \tau_\infty$, with $\tilde{W}$ defined previously. 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 the stated basic validity property of the bootstrap via Theorem 2 below; see Remark 3.10 for details.

3 Bootstrap validity under weak convergence to random distributions

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 distinguish between two concepts of bootstrap validity. The following definition employs the bootstrap p-value as a summary indicator of the accuracy of bootstrap inferences (see Remark 3.4 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 $\tau_n^* = \tau_n^*(D_n, W_n^*)$, where $D_n = D_n(X_n, E_n)$ denotes the data as a measurable function of some random elements X_n and E_n, not necessarily observable, whereas W_n^* are auxiliary variates used in the definition of the bootstrap procedure and defined jointly with (X_n, E_n) on a possibly expanded probability space. Let $p_n^* := P(\tau_n^* \le \tau_n|\, D_n)$ denote the bootstrap p-value.*

We say that bootstrap inference based on τ_n and τ_n^* is asymptotically valid conditionally on X_n if

$$P\left(p_n^* \leq q \middle| X_n\right) \xrightarrow{p} q \tag{14}$$

for all $q \in (0,1)$ as $n \to \infty$, so that bootstrap p-values are asymptotically $U(0,1)$ distributed conditionally on X_n.

We say that bootstrap inference based on τ_n and τ_n^* is asymptotically valid on average (over X_n) if

$$P\left(p_n^* \leq q\right) \to q \tag{15}$$

for all $q \in (0,1)$ as $n \to \infty$, so that bootstrap p-values are asymptotically $U(0,1)$ unconditionally.

Remark 3.1. In Definition 1, the original data D_n – hence the statistic τ_n – depend on two (possibly unobservable) components, one of which (X_n) the statistician might like to make inference conditional upon. Conversely, the bootstrap sample is defined in terms of the original data D_n and of some auxiliary variates collected in W_n^*. In standard applications, W_n^* are not D_n-meaurable and often are generated independently of D_n; the bootstrap statistic τ_n^* is not D_n-measurable either.

Remark 3.2. The examples from the previous section can be cast within the framework of Definition 1. For the regression models in Sections 2.1 and 2.3, we could consider $X_n := \{x_t\}_{t=1}^n$, $E_n := \{\varepsilon_t\}_{t=1}^n$ and $W_n^* = \{\varepsilon_t^*\}_{t=1}^n$, or alternatively, $X_n := x_{\lfloor n\cdot \rfloor}$, $E_n := n^{-1/2}\sum_{t=1}^{\lfloor n\cdot \rfloor}\varepsilon_t$ and $W_n^* := n^{-1/2}\sum_{t=1}^{\lfloor n\cdot \rfloor}\varepsilon_t^*$ if it is more convenient to deal with (convergent) random elements of a fixed space like $\mathscr{D}$. For the CUSUM test from Section 2.2 with original data $D_n := \{\varepsilon_t\}_{t=1}^n$, we can let $X_n := \{\varepsilon_{(t)}\}_{t=1}^n$ be the vector of order statistics associated to $\{\varepsilon_t\}_{t=1}^n$ and $E_n := \{\pi_t\}_{t=1}^n$ be the random permutation of $\{1,...,n\}$ (uniformly distributed conditionally on X_n) for which it holds that $\varepsilon_t = \varepsilon_{(\pi_t)}$ $(t = 1,...,n)$. Similarly, W_n^* is another random (uniformly distributed) permutation of $\{1,...,n\}$, independent of D_n.[4] Finally, for the Kolmogorov-Smirnov test from Section 2.4 we could have $X_n := x_{\lfloor n\cdot \rfloor}$, $E_n := n^{-1/2}\sum_{t=1}^n(\mathbb{I}_{\{\varepsilon_t \leq q(\cdot)\}} - (\cdot))$ and $W_n^* := n^{-1/2}\sum_{t=1}^n(\mathbb{I}_{\{\varepsilon_t^* \leq q(\cdot)\}} - (\cdot))$.

Remark 3.3. Asymptotic bootstrap validity conditionally on X_n implies validity on average, by bounded convergence. The converse does not hold.

Remark 3.4. The validity properties in Definition 1 ensure correct asymptotic size, conditionally on X_n or on average, 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 $P\left(\tau_n^* \leq \cdot | D_n\right)$ converges weakly in $\mathscr{D}(\mathbb{R})$ to a sample-path continuous random cdf (as we assume in this section), then correct asymptotic size is ensured also for bootstrap tests rejecting the null hypothesis when $\tilde{p}_n^* := P\left(\tau_n^* \geq \tau_n \middle| D_n\right) \leq \alpha$. Finally, if τ_n is not a genuine statistic but a function of a parameter evaluated at an unknown true value θ_0, assume that the condition $\tau_n(\theta_0) \in [\underline{q}, \overline{q}]$ can be solved for θ_0 as $\theta_0 \in S(D_n)$, where $\underline{q}, \overline{q}$ are D_n-measurable bootstrap quantiles satisfying $1 + P(\tau_n^* \leq \underline{q}|D_n) - P\left(\tau_n^* \leq \overline{q}|D_n\right) = \alpha$. Then the validity properties in Definition 1 and the weak convergence of $P\left(\tau_n^* \leq \cdot | D_n\right)$

[4]The CUSUM test based on residuals from a statistical model can be treated by including further components in E_n.

to a sample-path continuous random cdf ensure correct asymptotic coverage, conditionally on X_n or unconditionally, of $S(D_n)$ as a bootstrap confidence set for θ_0 for a confidence level of $1-\alpha$. □

3.1 Conditional bootstrap validity

We start by providing sufficient conditions for the bootstrap to be valid conditionally, i.e. in the sense of (14). The main requirement is the (joint) weak convergence of the distribution of the bootstrap statistic τ_n^*, conditional on the original data, and of a conditional distribution of the original statistic τ_n, to the same random limit distribution.

Theorem 1 *If, as $n \to \infty$, τ_n and τ_n^* of Definition 1 satisfy*

$$\left(P\left(\tau_n \leq \cdot \middle| X_n\right),\ P\left(\tau_n^* \leq \cdot \middle| D_n\right)\right) \xrightarrow{w} (F, F) \tag{16}$$

in $\mathscr{D}(\mathbb{R}) \times \mathscr{D}(\mathbb{R})$, where F is a sample-path continuous random cdf, then

$$\sup_{u \in \mathbb{R}} \left|P\left(\tau_n \leq u \middle| X_n\right) - P\left(\tau_n^* \leq u \middle| D_n\right)\right| \xrightarrow{p} 0 \tag{17}$$

as $n \to \infty$, and the bootstrap based on τ_n and τ_n^* is asymptotically valid conditionally on X_n as well as on average.

Proof. Convergence (17) is direct from (16), the assumed continuity of F and the behaviour of F at $\pm\infty$. Convergence (17) implies for $p_n^* = P\left(\tau_n^* \leq \tau_n \middle| D_n\right)$ that $p_n^* = F_n(u)|_{u=\tau_n} + o_p(1)$, where $F_n(u) := P\left(\tau_n \leq u \middle| X_n\right)$, $u \in \mathbb{R}$.[5] Let F_n^{-1} be the right-continuous generalized inverse of F_n, i.e., $F_n^{-1}(u) = \sup\{v \in \mathbb{R} : F_n(v) \leq u\}$. Then for $q \in (0,1)$ it holds that $\{F_n(u)|_{u=\tau_n} \leq q\} = \{\tau_n \leq F_n^{-1}(q)\}$ as an equality of events and

$$\begin{aligned}\left|P(F_n(u)|_{u=\tau_n} \leq q|X_n) - q\right| &= \left|P\left(\tau_n \leq F_n^{-1}(q)\middle| X_n\right) - q\right| = \left|F_n\left(F_n^{-1}(q)\right) - q\right| \\ &\leq \sup_{v \in \mathbb{R}} \left|F_n(v) - F_n(v-)\right| \xrightarrow{p} 0,\end{aligned}$$

the second equality by the X_n-measurability of $F_n^{-1}(q)$, and the zero limit from $F_n \xrightarrow{w} F$ in $\mathscr{D}(\mathbb{R})$ and the continuity of F. Thus, $F_n(u)|_{u=\tau_n}\big| X_n \xrightarrow{w}_p U(0,1)$. As $p_n^* = F_n(u)|_{u=\tau_n} + o_p(1)$ implies that $E(g(p_n^*)|X_n) = E(g(F_n(u)|_{u=\tau_n})|X_n) + o_p(1)$ for continuous and bounded real functions g, also $p_n^*|X_n \xrightarrow{w}_p U(0,1)$. Unconditional asymptotic uniformity of p_n^* follows by bounded convergence. □

The Gaussian regression of Section 2.1 is a rare case where a convergence like (16) can be established easily by a direct study of the involved finite-sample cdf's, see eq. (4). In practice it could be more natural to obtain (16) from a result about the weak convergence of the random measures induced by conditioning the distributions of τ_n and τ_n^*. In fact, by the argument of Daley and Vere-Jones (2008, pp.143-144), weak convergence of random measures implies weak convergence of the respective random cdf's in the Skorokhod topology. This observation gives rise to the next corollary.

[5] We write $F_n(u)|_{u=\tau_n}$ instead of $F_n(\tau_n)$ to avoid confusion with $P\left(\tau_n \leq \tau_n \middle| X_n\right)$.

Corollary 1 *Let D_n, X_n, E_n and W_n^* be as in Definition 1. Let there exist a random variable τ_∞ and a random element X_∞, not necessarily defined on the same probability space as D_n and W_n^*, such that*

$$(\tau_n|X_n, \tau_n^*|D_n) \xrightarrow{w}_w (\tau_\infty|X_\infty, \tau_\infty|X_\infty) \tag{18}$$

as $n \to \infty$, jointly in the sense of eq. (1). Let further $F(u) := P(\tau_\infty \leq u|X_\infty)$, $u \in \mathbb{R}$, define a random cdf with a.s. continuous sample paths. Then (16) holds, and the bootstrap based on τ_n and τ_n^* is asymptotically valid conditionally on X_n as well as on average.

Remark 3.5. [Examples, cont'd] For the CUSUM bootstrap test of Section 2.2, $\tau_\infty = \rho_\alpha(S)$, $X_\infty = S$ and condition (18) takes the form of (5) or (6), whereas $F(u) = P(\rho_\alpha(S) \leq u|S)$. Asymptotic validity of the bootstrap in a conditional sense will follow from Corollary 1 if F has a.s. continuous sample paths. For the co-integrating regression example, under the extra assumptions of Section 2.3(ii), we can consider $\tau_\infty = (\int B_\eta^2)^{-1} \int B_\eta dB_\varepsilon$, $X_\infty = B_\eta$ and condition (18) holds in the form of (9), with F having Gaussian cdf's as sample paths, which are continuous. By Corollary 1, this implies asymptotic bootstrap validity conditional on the regressors. For the goodness of fit test, the relevant objects could be $\tau_\infty = \sup_{u\in[0,1]} |\tilde{W}(u)|$ and X_∞ chosen as either (M, ξ) or the weak limit of $n^{-\alpha/2}x_{\lfloor n\cdot\rfloor}$ in $\mathscr{D}$. The sample paths of $P(\tau_\infty \leq \cdot|M, \xi)$ are a.s. continuous, e.g., by Proposition 3.2 of Linde (1989) applied conditionally on M, ξ. In order to guarantee that eq. (18) holds, conditions restricting the dependence between regressors and disturbances would be needed; indpendence is certainly sufficient but it is beyond the scope of this paper to explore possible generalizations. □

We conclude this section with a proof of two properties of the bootstrap when (16) does not hold. First, we show that bootstrap validity conditionally on X_n is lost when τ_n given X_n and τ_n^* given the data converge to distinct random limits; the discussion in Section 2.3(iv) falls within the framework of this result. Second, we establish that the bootstrap may be valid at least on average, provided that the limit distribution of the bootstrap statistic is a conditional average of the limit distribution of τ_n given X_n. We make the simplifying assumption that τ_n^* depends on the data D_n through X_n alone, which involves no loss of generality in the (counterexemplary) negative part of the next proposition. The positive part of the proposition will be extended in Theorem 2 below with no recourse to this simplification.

Proposition 1 *With the notation of Definition 1, assume that $P(\tau_n^* \leq \cdot|\, D_n) = P(\tau_n^* \leq \cdot|\, X_n)$. Let it hold, as $n \to \infty$, that*

$$(P(\tau_n \leq \cdot|\, X_n),\ P(\tau_n^* \leq \cdot|\, X_n)) \xrightarrow{w} (F, G) \tag{19}$$

in $\mathscr{D}(\mathbb{R}) \times \mathscr{D}(\mathbb{R})$, where F and G are random cdf's with a.s. continuous sample paths. Let G^{-1} be the right-continuous generalized inverse of G. Then, as $n \to \infty$:

(i) the bootstrap p-value $p_n^* = P(\tau_n^* \leq \tau_n|\, D_n)$ satisfies

$$P(p_n^* \leq q|\, X_n) \xrightarrow{w} F\left(G^{-1}(q)\right) \tag{20}$$

for almost all $q \in (0,1)$, so the bootstrap based on τ_n and τ_n^* is asymptotically valid conditionally on X_n iff $F = G$;

(ii) provided that $G(u) = E(F(u)|\mathcal{G}_\infty)$, $u \in \mathbb{R}$, for some sub-σ-algebra $\mathcal{G}_\infty$ on the probability space where F and G are defined, the bootstrap based on τ_n and τ_n^* is asymptotically valid on average.

REMARK 3.6. An instance of (20) with $F \neq G$ is provided in Section 2.3(iv), where

$$F(u) = P(M^{-1/2}(\sqrt{\omega_{\varepsilon|\eta}}\xi_1 + \sqrt{1-\omega_{\varepsilon|\eta}}\xi_2) \leq u|M, \xi_2) = \Phi(\omega_{\varepsilon|\eta}^{-1/2}\{M^{1/2}u - \sqrt{1-\omega_{\varepsilon|\eta}}\xi_2\}),$$
$$G(u) = P(M^{-1/2}(\sqrt{\omega_{\varepsilon|\eta}}\xi_1 + \sqrt{1-\omega_{\varepsilon|\eta}}\xi_2) \leq u|M) = \Phi(M^{1/2}u),\ u\in \mathbb{R}.$$

Thus, $F(G^{-1}(q)) = \Phi(\omega_{\varepsilon|\eta}^{-1/2}\{\Phi^{-1}(q) - \sqrt{1-\omega_{\varepsilon|\eta}}\xi_2\})$ is the random cdf corresponding to the limit in eq. (11). Still, as it holds that $G(u) = E(F(u)|\mathcal{G}_\infty)$, $u \in \mathbb{R}$, for $\mathcal{G}_\infty = \sigma(M)$, Proposition 1 implies that the bootstrap is asymptoticly validit on average, as was directly concluded in Section 2.3(i) by exploiting the simple construction of the bootstrap. □

3.2 Bootstrap validity on average

We proceed with a general result about 'on average' asymptotic validity of the bootstrap under weaker requirements than in Theorem 1. Specifically, assumption (16) is relaxed to the requirement that the unconditional limit distribution of the original statistic τ_n should be an average of the random limit distribution of the bootstrap statistic τ_n^* given the data. Regarding the scope and the ease of application, this requirement has the advantage to not be based on a conditional analysis of τ_n. The cost is the loss of convergence (17), of conditional bootstrap validity, and hence, of the bootstrap as a tool for conditional inference.

Theorem 2 *Let there exist a rv τ_∞ and a random element X_∞ of a Polish space $\mathcal{S}$, both defined on the same probability space, such that $(\tau_n, G_n) \xrightarrow{w} (\tau_\infty, G)$ in $\mathbb{R} \times \mathscr{D}(\mathbb{R})$ for $G_n(u) := P(\tau_n^* \leq u|D_n)$ and $G(u) := P(\tau_\infty \leq u|X_\infty)$, $u \in \mathbb{R}$. If the data D_n ($n \in \mathbb{N}$) is a random element of a Polish space and if the random cdf G is sample-path continuous, then the bootstrap based on τ_n and τ_n^* is asymptotically valid on average.*

PROOF OF THEOREM 2. The random element (τ_∞, G) of $\mathbb{R}\times\mathscr{D}(\mathbb{R})$ is a measurable function of (τ_∞, X_∞) and this function is fully determined 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 proability space where $(\tau_n, G_n) \to (\tau_\infty, G)$ a.s. in $\mathbb{R} \times \mathscr{D}(\mathbb{R})$ and $G(\cdot) = P(\tau_\infty \leq \cdot|X_\infty)$ still holds. We show that $G_n(\tau_n) \xrightarrow{a.s.} U(0,1)$ on the special probability space, so in general $G_n(\tau_n) \xrightarrow{w} U(0,1)$.

Since G is continuous and G_n, G are (random) cdf's, $G_n \xrightarrow{a.s.} G$ in $\mathscr{D}(\mathbb{R})$ implies that $\sup_{u\in\mathbb{R}} |G_n(u) - G(u)| \xrightarrow{a.s.} 0$. Therefore, $G_n(\tau_n) - G(u)|_{u=\tau_n} \xrightarrow{a.s.} 0$. Since $\tau_n \xrightarrow{a.s.} \tau_\infty$ and G is uniformly continuous, it holds further that $G(u)|_{\tau_n} - G(u)|_{\tau_\infty} \xrightarrow{a.s.} 0$, so also $G_n(\tau_n) - G(u)|_{u=\tau_\infty} \xrightarrow{a.s.} 0$. From the equality of events $\{G(u) \leq q\} = \{u \leq G^{-1}(q)\}$, $q \in (0,1)$, it follows that

$$P\left(G(u)|_{u=\tau_\infty} \leq q|\, X_\infty\right) = P\left(\tau_\infty \leq G^{-1}(q)\middle|\, X_\infty\right) = G\left(G^{-1}(q)\right) = q,$$

the penultimate equality because $G^{-1}(q)$ is X_∞ -measurable. We conclude that $G(u)|_{u=\tau_\infty}$ is uniformly distributed conditionally on X_∞, and hence, unconditionally. Then $G_n(\tau_n) \overset{a.s.}{\to} G(u)|_{u=\tau_\infty} = U(0,1)$. □

Remark 3.7. A trivial special case of Theorem 2 is that where τ_∞ and X_∞ are independent. In this case the bootstrap distribution of τ_n^* estimates consistently the limiting unconditional distribution of τ_n and the bootstrap is asymptotically valid in the usual sense.

Remark 3.8. Proposition 1(ii) is another special case of Theorem 2. Indeed, on an extension of the probability space where F and G are defined, there exists a rv $\theta \sim U(0,1)$ independent of F. Then (19) implies that $(\tau_n, G_n) \overset{w}{\to} (\tau_\infty, G)$ for $\tau_\infty := F^{-1}(\theta)$, $X_\infty := F$ and $\mathcal{S} := \mathscr{D}(\mathbb{R})$.

Remark 3.9. A third special case of Theorem 2 involves stable convergence of the original statistic τ_n (see Häusler and Luschgy, 2015, for a definition). With the notation of Theorem 2, 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 $\tau_n \to \tau_\infty$ stably and $G_n \overset{p}{\to} G$. Then $(\tau_n, G_n) \overset{w}{\to} (\tau_\infty, G)$ by Theorem 3.7(b) of Häusler and Luschgy (2015). For instance, in the financial econometric literature on integrated volatility, a result of the form $\tau_n \to \tau_\infty$ is contained in Theorem 3.1 of Jacod *et al.* (2009) for τ_n defined as a t-like statistic for integrated volatility, whereas the corresponding $G_n \overset{p}{\to} G$ result is established in Theorem 3.1 of Hounyo *et al.* (2017) for a combined wild and blocks-of-blocks bootstrap introduced in the latter paper.

Remark 3.10. [Examples, cont'd] Return to the goodness-of-fit test example of Section 2.4. As proved in Section 6, it holds that $(\tau_n, X_n, G_n) \overset{w}{\to} (\tau_\infty, X_\infty, G)$ with $X_n := n^{-\alpha/2}x_{\lfloor n\cdot\rfloor} \in \mathscr{D}$ and $\tau_\infty := \sup_{u\in[0,1]} |\tilde{W}(u)|$. As G is sample-path continuous by Proposition 3.2 of Linde (1989), Theorem 2 guarantess on average asymptotic validity of the bootstrap. This conclusion holds without extra conditions restricting the dependence between regressors and disturbances; cf. Remark 3.5.

Remark 3.11. Even if, as in the previous remark, there exist random elements $X_n \in \mathcal{S}$ measurable with respect to the data and such that the condition of Theorem 2 holds in the stronger form $(\tau_n, X_n, G_n) \overset{w}{\to} (\tau_\infty, X_\infty, G)$, it need not hold that $\tau_n|X_n \overset{w}{\to}_w \tau_\infty|X_\infty$. Still, by extended Skorokhod coupling, if the data and (τ_∞, X_∞) are redefined on a special probability space where $(\tau_n, X_n, G_n) \overset{a.s.}{\to} (\tau_\infty, X_\infty, G)$, then the convergence $\tau_n \overset{a.s.}{\to} \tau_\infty$ implies, by the martingale convergence theorem (Loève, 1978, p.75, ex.10), that

$$\tau_n|\{X_1, ..., X_n\} \overset{w}{\to}_{a.s.} \tau_\infty|\mathcal{F}_\infty,$$

where $\mathcal{F}_\infty$ is the null-sets completion of $\sigma(\{X_n\}_{n\in\mathbb{N}})$.[6] From the completeness of $\mathcal{F}_\infty$, it follows that X_∞ is $\mathcal{F}_\infty$-measurable as the a.s. limit of $\mathcal{F}_\infty$-measurable functions. Hence, for $F(u) := P(\tau_\infty \le u|\mathcal{F}_\infty)$, we find that $G(u) = E[F(u)|X_\infty]$, $u \in \mathbb{R}$. Thus, implicitly, the limit bootstrap distribution is a mixture of limiting conditional distributions of the original statistic (upon a redefinition of the probability space), similarly to Proposition 1(ii). □

[6]Other conditioning choices are possible, e.g. $\sigma\{X_n, X_{n+1}, ...\}$ on the left side and the completion of the tail σ-algebra of $\{X_n\}$ on the right side.

4 Bootstrap tests of parameter constancy

4.1 Problem and standard asymptotic theory

Here we apply the results of Section 3 to the classical problem of parameter constancy testing in regression models (Chow, 1960; Quandt, 1960; Nyblom, 1989; Andrews, 1993; Andrews and Ploberger, 1994). Specifically, we deal with bootstrap implementations when the moments of the regressors may be unstable over time; see Hansen (2000) and Zhang and Wu (2012), inter alia.

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

$$y_{nt} = \beta_t' x_{nt} + \varepsilon_{nt} \qquad (t = 1, 2, ..., n). \tag{21}$$

The null hypothesis of parameter constancy is $\mathsf{H}_0 : \beta_t = \beta_1 \, (t = 2, ..., n)$, which is tested here agains the alternative $\mathsf{H}_1 : \beta_t = \beta_1 + \theta_n \mathbb{I}_{\{t \geq n^*\}} \; (t = 2, ..., n)$, where $n^* := \lfloor r^* n \rfloor$ and $\theta_n \neq 0$ respectively denote the timing and the magnitude of the possible break,[7] both assumed unknown to the statistician. The so-called break fraction r^* 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 (Quandt, 1960; Andrews, 1993),[8] 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 $\mathcal{H}$

(i) *(mda) ε_{nt} is a martingale difference array (mda) 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 $n^{-1} \sum_{t=1}^{\lfloor nr \rfloor} \varepsilon_{nt}^2 \xrightarrow{p} r(E\varepsilon_{nt}^2) = r\sigma^2 > 0$ as $n \to \infty$, for all $r \in (0, 1]$;*

(iii) *(non-stationarity) in $\mathscr{D}_{m \times m} \times \mathscr{D}_{m \times m} \times \mathscr{D}_m$:*

$$\left(\frac{1}{n} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_{nt} x_{nt}', \frac{1}{n\sigma^2} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_{nt} x_{nt}' \varepsilon_t^2, \frac{1}{n^{1/2}\sigma} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_{nt} \varepsilon_{nt} \right) \xrightarrow{w} (M, V, N)$$

as $n \to \infty$, 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 $E\{N(r_1) N(r_2)'\} = V(r_1)$ $(0 \leq r_1 \leq r_2 \leq 1)$.

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

[8] Alternative tests can be considered similarly, see e.g. Hansen (2000, sec.2).

REMARK 4.1. A special case of Assumption $\mathcal{H}$ is obtained when the regressors satisfy the weak convergence $x_{n\lfloor n\cdot\rfloor} \overset{w}{\to} U(\cdot)$ in $\mathscr{D}_m$ (see, e.g., the example from Section 2.3). In this case, $M(\cdot) = \int_0^{\cdot} UU'$ and, under appropriate restrictions on the squared errors, V equals M (this equality will be important for the asymptotic validity of the first bootstrap procedure discussed below). For instance, if $\sup_n \sup_{t=1,\dots,n} E|E(\varepsilon_{nt}^2 - \sigma^2|\mathcal{F}_{n,t-m})| \to 0$ as $m \to \infty$ for some filtrations $\mathcal{F}_{n,t}$, $n \in \mathbb{N}$, to which ε_{nt}^2 are adapted, then $M = V$ (see Theorem A.1 of Cavaliere and Taylor, 2009).

REMARK 4.2. In a recent paper Zhang and Wu (2012) assume that the x_{nt}'s are 'locally stationary', i.e. of the form $x_{nt} = G(t/n; \mathcal{F}_t)$ for $\mathcal{F}_t := \sigma(\dots, \varepsilon_{t-1}, \varepsilon_t)$, $\{\varepsilon_t\}$ being a sequence of i.i.d. rv's and G a sufficiently smooth random function. Regressors of this form are not ruled out by Assumption $\mathcal{H}$. The main difference is that Zhang and Wu (2012) place (smoothness and moment) restrictions on the function G, whereas Assumption $\mathcal{H}$ restricts the large sample behaviour of the sample moments.[9]

REMARK 4.3. The case of (asymptotically) covariance stationary regressors is covered by Assumption $\mathcal{H}$ and correspponds to $M(r) = r\Sigma$ for a non-random variance matrix Σ. $\square$

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

$$\mathscr{F}_n \overset{w}{\to} \sup_{r\in[\underline{r},\overline{r}]} \left\{ \tilde{N}(r)' \tilde{M}(r)^{-1} \tilde{N}(r) \right\} \tag{22}$$

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 Bessell process; see Andrews (1993). In the general case, however, since the asymptotic distribution in (22) depends on the joint distribution of the limiting processes M, N, V, which is unspecified under Assumption $\mathcal{H}$, asymptotic inference based on (22) is unfeasible. Simulation methods as the bootstrap can therefore be appealing devices to compute p-values associated with $\mathscr{F}_n$. In particular, if the statistician is not interested in modeling the distribution of the regressors x_{nt}, but opts instead for inference *conditional* on x_{nt}, it appears natural to resort to the fixed regressor bootstrap, where x_{nt}, $t = 1, \dots, n$, are fixed across bootstrap repetitions.

4.2 Bootstrap implementations and (random) limit theory

Following Hansen (2000), we discuss here two implementations of the fixed-regressor bootstrap. In the first procedure, the bootstrap sample is generated as $y_{t,1}^* = w_t^*$ $(t = 1, \dots, n)$, with w_t^* i.i.d. $N(0,1)$ and independent of the original data $D_n := \{x_{nt}, y_{nt}\}_{t=1}^n$. The bootstrap statistic is then defined as

$$\mathscr{F}_{1,n}^* := \max_{r\in[\underline{r},\overline{r}]} F_{1,\lfloor nr\rfloor}^*,$$

[9]Zhang and Wu (2012) propose, without theory, a fully parametric bootstrap for computing p-values associated to their parameter constancy test. We conjecture that validity of the bootstrap in their framework could be established using the general theory of the present paper.

where $F^*_{1,\lfloor nr\rfloor}$ is the usual F statistic for testing the auxiliary null hypothesis $\theta^* = 0$ in the regression

$$y^*_{t,1} = \beta^{*\prime} x_{nt} + \theta^{*\prime} x_{nt}\mathbb{I}_{\{t\geq\lfloor rn\rfloor\}} + \text{error}^*_{nt}. \tag{23}$$

The second bootstrap procedure is a fixed-regressor wild bootstrap (Wu, 1986; Liu, 1988; Mammen, 1993) and is introduced to accommodate possible conditional heteroskedasticity of ε_{nt}.[10] It is based on the residuals $\tilde{e}_{nt}$ from the OLS regression of y_{nt} on x_{nt} and $x_{nt}\mathbb{I}_{\{t\geq\lfloor\tilde{r}n\rfloor\}}$, 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

$$\mathscr{F}^*_{2,n} := \max_{r\in[\underline{r},\bar{r}]} F^*_{2,\lfloor nr\rfloor},$$

where $F^*_{2,\lfloor nr\rfloor}$ is the F statistic for the auxiliary null that $\theta^* = 0$ in (23), with $y^*_{t,1}$ now replaced by the wild bootstrap innovation $y^*_{t,2} := \tilde{e}_{nt} w^*_t$.

The weak limit properties of the bootstrap statistics $\mathscr{F}^*_{1,n}$ and $\mathscr{F}^*_{2,n}$ are stated in the next theorem.

Theorem 3 *Under Assumption $\mathcal{H}$, and for $i = 2$, additionally, under H_0, it holds that*

$$\mathscr{F}^*_{i,n} \overset{w^*}{\to}_w \sup_{r\in[\underline{r},\bar{r}]} \left\{\tilde{N}_i(r)'\tilde{M}(r)^{-1}\tilde{N}_i(r)\right\}\Bigg| M, V,\ i = 1, 2, \tag{24}$$

where $\tilde{M}(r) := M(r) - M(r)M(1)^{-1}M(r)$, $\tilde{N}_i(r) := N_i(r) - M(r)M(1)^{-1}N_i(1)$ ($i = 1, 2$), N_1 conditionally on M, V is a zero-mean Gaussian process with covariance kernel $E\{N_1(r_1)N_1(r_2)'|M\} = M(r_1)$ for $\underline{r} \leq r_1 \leq r_2 \leq \bar{r}$, and $N_2|M, V$ is distributed like $N|M, V$ of Assumption $\mathcal{H}$.

Remark 4.4. Theorem 3 establishes that, in general, the weak limits of the fixed-regressor bootstrap statistics are *random*. In particular, they are distinct from the limit in eq. (22) 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 Theorems 5 and 6 of Hansen, 2000). To highlight the source of the limiting randomness, consider the case $M = V$. In this case, for fixed $r \in [\underline{r}, \bar{r}]$, $\tilde{N}_i(r)'\tilde{M}(r)^{-1}\tilde{N}_i(r)$ is $\chi^2(m)$-distributed conditionally on M ($i = 1, 2$), and hence, is independent of M. Nevertheless, the non-contemporaneous covariances of the process $\tilde{M}(r)^{-1/2}\tilde{N}_i(r)$ conditional on M do depend on M, and therefore, the limit distributions in (24) are random whenever M is random.

Remark 4.5. In addition to the previous remark, consider the case of a scalar regressor $x_{nt} \in \mathbb{R}$. By a change of variable (as in Theorem 3 of Hansen, 2000), convergence (24) reduces to

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

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. □

Next, we formulate conditions under which the fixed-regressor bootstrap is asymptotically valid in the sense of Definition 1.

[10] It can also accommodate forms of unconditional heteroskedasticity that violate Assumption $\mathcal{H}$(ii). Following Hansen (2000), we do not pursue this extension.

4.3 Bootstrap validity

Although the bootstrap procedures do not mimic the asymptotic (unconditional) distribution in (22), bootstrap validity on average (over $X_n := \{x_{nt}\}_{t=1}^n$) can be established using Theorem 2. For establishing bootstrap validity conditionally on X_n by means of Theorem 1, we strengthen Assumption $\mathcal{H}$ as follows.

ASSUMPTION $\mathcal{C}$. *Assumption $\mathcal{H}$ holds and, as random measures on $\mathscr{D}_{m\times m} \times \mathscr{D}_{m\times m} \times \mathscr{D}_m$,*

$$\left(\frac{1}{n}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}', \frac{1}{n\sigma^2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}'\varepsilon_{nt}^2, \frac{1}{n^{1/2}\sigma}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}\varepsilon_{nt}\right)\Bigg| X_n \xrightarrow{w}_w (M, V, N)\,|M, V\ .$$

REMARK 4.6. 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 example in Section 2.3(iv) illustrates the effect of dependence on conditional limits. With the notation of Section 2.3(iv), let $x_{nt} := n^{-1/2}x_t$ and $\varepsilon_{nt} := \varepsilon_t$. Then Assumption $\mathcal{H}$(iii) holds with $M = V = \int_0^\cdot B_\eta^2$ and $N = \sqrt{\omega_{\varepsilon|\eta}}\int_0^\cdot B_\eta dB_{y1} + \sqrt{1-\omega_{\varepsilon|\eta}}\int_0^\cdot B_\eta dB_{y2}$. However, $n^{-1/2}\sum_{t=1}^n x_{nt}\varepsilon_{nt}\big| X_n \xrightarrow{w}_w N(1)|\{M, V, \int B_\eta dB_{y2}\}$ with the limit distributed differently from $N(1)|M, V$. Thus, the actual conditioning in the limit is more complex than required in Assumption $\mathcal{C}$. □

We now establish the asymptotic validity of the bootstrap parameter constancy tests under the introduced assumptions.

Theorem 4 . *Let the parameter constancy hypothesis H_0 hold for model (21). Then, under Assumption $\mathcal{H}$, the bootstrap based on $\tau_n = \mathscr{F}_n$ and $\tau_n^* = \mathscr{F}_{2,n}^*$ is asymptotically valid on average. If Assumption $\mathcal{C}$ holds, then (17) holds for $X_n = \{x_{nt}\}_{t=1}^n$ and the bootstrap based on $\mathscr{F}_n$ and $\mathscr{F}_{2,n}^*$ is asymptotically valid also conditionally on X_n. The same results hold for the bootstrap based on $\tau_n = \mathscr{F}_n$ and $\tau_n^* = \mathscr{F}_{1,n}^*$ provided that $M = V$.*

5 Conclusions

In this study of bootstrap statistics with random limit distributions given the data, we have provided a formal analysis of the asymptotic validity of bootstrap inference, in a conditional and on-average sense. For both types of asymptotic validity, we have established sufficient conditions. These differ mainly in their demands on the dependence structure of the data, and are more restrictive for conditional validity to hold. We have seen that this difference is essential and not an artefact of our approach.

6 Additional proofs

PROOF OF EQ. (10). Let $\sum_{s=1}^t \varepsilon_s = \mathring{Y}_t + Y_t^e$ with $\mathring{Y}_t := \sum_{s=1}^t\{\varepsilon_s - E(\varepsilon_s|\eta_s)\}$ and $Y_t^e := \sum_{s=1}^t E(\varepsilon_s|\eta_s)$. A standard fact is that $n^{-1/2}\left(\mathring{Y}_{\lfloor n\cdot\rfloor}, Y_{\lfloor n\cdot\rfloor}^e, \sum_{s=1}^{\lfloor n\cdot\rfloor}\eta_s\right) \xrightarrow{w} (\sqrt{\omega_{\varepsilon|\eta}}B_{y1}, \sqrt{1-\omega_{\varepsilon|\eta}}B_{y2}, B_x)$

in $\mathscr{D}_3$, where (B_{y1}, B_{y2}, B_x) is a standard Brownian motion in $\mathbb{R}^3$. Further, by the conditional invariance principle of Rubstein (1996),

$$n^{-1/2}\,\mathring{Y}_{\lfloor n\cdot\rfloor}\Big|\sum_{s=1}^{\lfloor n\cdot\rfloor}\eta_s \xrightarrow{w}_p \sqrt{\omega_{\varepsilon|\eta}}B_{y1} \overset{d}{=} \sqrt{\omega_{\varepsilon|\eta}}B_{y1}\Big|\, B_{y2}, B_x$$

as a convergence of random measures on $\mathscr{D}$. Since $\sigma(\sum_{s=1}^{\lfloor n\cdot\rfloor}\eta_s) = \sigma(Y^e_{\lfloor n\cdot\rfloor}, \sum_{s=1}^{\lfloor n\cdot\rfloor}\eta_s) = \sigma(\{x_t\}_{t=1}^n)$, the convergence

$$n^{-1/2}\left(\mathring{Y}_{\lfloor n\cdot\rfloor}, Y^e_{\lfloor n\cdot\rfloor}, \sum_{s=1}^{\lfloor n\cdot\rfloor}\eta_s\right)\Bigg|\{x_t\}_{t=1}^n \xrightarrow{w}_w \left(\sqrt{\omega_{\varepsilon|\eta}}B_{y1}, \sqrt{1-\omega_{\varepsilon|\eta}}B_{y2}, B_x\right)\Big|\, B_{y2}, B_x$$

follows by Theorem 2.1 of Crimaldi and Pratelli (2005), for random measures on $\mathscr{D}_3$. By using conditional convergence to stochastic integrals (Theorem 5 of Georgiev *et al.*, 2016) for the statistic τ_n of (7), eq. (10) follows. □

Proof of Eq. (12) and Remark 3.10. By extended Skorokhod coupling (Corollary 5.12 of Kallenberg (1997)), we can regard the data as defined on a special probability space such that $(n^{-\alpha/2}x_{\lfloor n\cdot\rfloor}, n^{-1/2}\sum_{t=1}^n\{\mathbb{I}_{\{\varepsilon_t\le q(\cdot)\}} - u\}) \to (X_\infty, W)$ a.s. in $\mathscr{D}\times\mathscr{D}$; then, by a product-space construction, we can expand this space to define also an i.i.d. standard Gaussian sequence ε_t^*. Consider outcomes in the component-space of $n^{-\alpha/2}x_{\lfloor n\cdot\rfloor}$ such that $(n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n) \to (M,\xi)$, $n^{-(\alpha+1)/2}\sup|x_{\lfloor n\cdot\rfloor}| \to 0$ and $M > 0$; such outcomes are almost all. For every such outcome, $(n^{1/2}W_n^*, M_n^{1/2}\hat{\beta}^*)$ is tight in $\mathscr{D}\times\mathbb{R}$ because $n^{1/2}W_n^*$ and $M_n^{1/2}\hat{\beta}^*$ are tight in $\mathscr{D}$ and $\mathbb{R}$ resp., and its finite-dimensional distributions converge to those of (W,b) by the Lyapunov CLT. It follows that $(n^{1/2}W_n^*, M_n^{1/2}\hat{\beta}^*)|x_{\lfloor n\cdot\rfloor} \xrightarrow{w}_{a.s.} (W,b)$, and further, that $(n^{1/2}W_n^*, M_n^{1/2}\hat{\beta}^*, n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n)|x_{\lfloor n\cdot\rfloor} \xrightarrow{w}_{a.s.} (W,b,M,\xi)|M,\xi$, by the $x_{\lfloor n\cdot\rfloor}$-measurability of $(n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n)$. Still further, , by continuity considerations, $(n^{1/2}W_n^*,\, n^{(\alpha+1)/2}\hat{\beta}^*,\, n^{-\alpha/2-1}\xi_n)|x_{\lfloor n\cdot\rfloor} \xrightarrow{w}_{a.s.} \left(W, M^{-1/2}b, \xi\right)\big|\, M,\xi$ on the special probability space. This implies (12) on a general probability space. Back on the special probability space, we conclude by the same argument as for eq. (13) that $\tau_n^*|x_{\lfloor n\cdot\rfloor} \xrightarrow{w}_p \tau_\infty|X_\infty$ so that $G_n(\cdot) := P(\tau_n^* \le \cdot|D_n) \xrightarrow{p} G(\cdot) := P(\tau_\infty \le \cdot|X_\infty)$ in $\mathscr{D}(\mathbb{R})$. As further $\tau_n \xrightarrow{p} \tau_\infty$ on the special probability space, we can collect the previous convergence facts into $(\tau_n, X_n, n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n, G_n) \xrightarrow{p} (\tau_\infty, X_\infty, M, \xi, G)$ on that same space, which proves that $(\tau_n, Y_n, G_n) \xrightarrow{w} (\tau_\infty, Y_\infty, G)$ on a general probability space, for the two choices of Y_n given in Remark 3.10.

Proof of Proposition 1. Let $F_n(u) := P(\tau_n \le u|\, X_n)$ and $G_n(u) := P(\tau_n^* \le u|\, X_n)$ for $u\in\mathbb{R}$, such that, by (19), $(F_n, G_n) \xrightarrow{w} (F,G)$ in $\mathscr{D}(\mathbb{R})\times\mathscr{D}(\mathbb{R})$. By assumption, $p_n^* = G_n(\tau_n)$. If G_n^{-1} stands for the right-continuous generalized inverse of G_n, it follows from the cmt that $(F_n, G_n^{-1}(q)) \xrightarrow{w} (F, G^{-1}(q))$ in $\mathscr{D}(\mathbb{R})\times\mathbb{R}$ for every $q\in(0,1)$ at which G^{-1} is a.s. continuous. For such q we find that

$$P(p_n^* \le q|X_n) = P(\tau_n \le G_n^{-1}(q)|X_n) = F_n(G_n^{-1}(q)) \xrightarrow{w} F\left(G^{-1}(q)\right), \tag{25}$$

the second equality by the X_n-measurability of $G_n^{-1}(q)$, and the limit by the cmt and the continuity of F. Since such $q \in (0,1)$ are all but countably many, (20) follows. Asymptotic validity of the bootstrap conditional on X_n requires that $F(G^{-1}(q)) = q$ for almost all $q \in (0,1)$, which by the continuity of F and G reduces to $F = G$. For part (ii), let $g(\cdot) = \min\{\cdot, 1\}\mathbb{I}_{[0,\infty)}(\cdot)$. By the definition of weak convergence, (25) implies

$$\begin{aligned} P(p_n^* \leq q) &= E\{g(P(p_n^* \leq q|X_n))\} \overset{w}{\to} E\{F(G^{-1}(q))\} \\ &= E\{E[F(G^{-1}(q))|\mathcal{G}_\infty\} = E\{G(G^{-1}(q))\} = q \end{aligned}$$

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

7 References

Andrews, D.W.K. (1997): A conditional Kolmogorov test, *Econometrica* 65, 1097–1128.

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

Andrews, D.W.K. (1999): Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space, *Econometrica* 68, 399–405.

Andrews, D.W.K. and W. Ploberger (1994): Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383–1414.

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

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

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.

Birnbaum A. (1962): On the foundations of statistical inference, *Journal of the American Statistical Association* 57, 269-.306.

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., I. Georgiev and A.M.R. Taylor (2016): Sieve-based inference for infinite-variance linear processes, *Annals of Statistics* 44, 1467-1494.

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

Chow G. (1960): Tests of equality between sets of coefficients in two linear regressions, *Econometrica* 28, 591–605.

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

Daley D. and D. Vere-Jones (2008): *An Introduction to the Theory of Point Processes*, Volume II: General Theory and Structure, Springer-Verlag.

Georgiev I., Harvey D., Leybourne S., Taylor A.M.R. (2016): A bootstrap stationarity test for predictive regression invalidity. Mimeo.

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

Hansen B. (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.

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

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

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

Kallenberg O. (1987): *Foundations of Modern Probability*, Springer-Verlag: Berlin

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

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

Liu R.Y. (1988): Bootstrap procedures under some non i.i.d. models, *Annals of Statistics* 16, 1696–1708.

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

Loève M. (1978): *Probability Theory II*, Springer-Verlag: Berlin

Mammen E. (1993): Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics* 21, 255–285.

NYBLOM J., (1989): Testing for the constancy of parameters over time, *Journal of the American Statistical Association* 84, 223-230.

PLOBERGER, W. AND W. KRÄMER (1992): The CUSUM test with OLS residuals, *Econometrica* 60, 271–285.

QUANDT R. (1960): Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of the American Statistical Association* 55, 324-330.

RUBSTEIN 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.

SAIKKONEN, P. (1991): Asymptotically Efficient Estimation of Cointegration Regressions, *Econometric Theory* 7, 1–21.

WU, C.F.J. (1986): Jackknife, bootstrap, and other resampling methods, *Annals of Statistics* 14, 1261–1295.

ZHANG T. AND W. WU (2012): Inference of time-varying regression models, *Annals of Statistics* 40, 1376-1402.

SUPPLEMENT: PROOFS TO THE PARAMETER CONSTANCY SECTION

Proof of Theorem 3. Let $(M_n, \tilde{V}_n) := (n^{-1}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}', n^{-1}\sigma^{-2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}'\tilde{e}_{nt}^2)$. As $\tilde{V}_n = n^{-1}\sigma^{-2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x_{nt}'\varepsilon_{nt}^2 + o_p(1)$ under H_0, under Assumption $\mathcal{H}$ it holds that $(M_n, \tilde{V}_n) \xrightarrow{w} (M, V)$ in $\mathscr{D}_{m\times m}\times\mathscr{D}_{m\times m}$. The original data $D_n := \{x_{nt}, y_{nt}\}_{t=1}^n$ and the bootstrap multipliers $\{w_t^*\}_{t\in\mathbb{N}}$ can be regarded (upon padding with zeroes) as defined on the Polish space $(\mathbb{R}^\infty)^{k+2}$. Therefore, by Corollary 5.12 of Kallenberg (1997), there exists a special probability space where (M, V), and for every $n \in \mathbb{N}$, also the original data and the bootstrap data can be redefined, maintaining their distribution (with a slight abuse, we also maintain the notation), such that $(M_n, \tilde{V}_n) \overset{a.s.}{\to} (M, V)$.

The first ingredient of the proof is the conditional convergence

$$N_i^* | \, M_n, \tilde{V}_n \xrightarrow{w}_{a.s.} N_i | M, V \tag{26}$$

on the special probability space, for $N_i^* := n^{-1/2}\sigma^{-1}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}y_{t,i}^*$, as an a.s. convergence of random measures on $\mathscr{D}_m$. As N_i^* conditional on the data is a zero-mean Gaussian process with independent increments and variance function $\sigma^{-2}M_n$ and $\tilde{V}_n$ resp. for $i = 1, 2$, the argument for Theorem 5 of Hansen (2000) yields (26). We notice that, for $i = 1$, the process N_i^* conditionally on M_n is independent of $\tilde{V}_n$, and therefore, for $i = 1$, (26) and the subsequent results hold independently of whether H_0 is true or not (which plays a role in the asymptoics of $\tilde{V}_n$).

The second ingredient of the proof is the joint convergence

$$\left(M_n, \tilde{V}_n, N_i^*\right) \xrightarrow{w} (M, V, N_i) \tag{27}$$

in $(\mathscr{D}_{m\times m})^2\times\mathscr{D}_m$, which follows from (26) and the marginal convergence $(M_n, \tilde{V}_n) \overset{a.s.}{\to} (M, V)$ in $(\mathscr{D}_{m\times m})^2$. Jointly (26) and (27), by Corollary 4.1 and Proposition 4.3 of Crimaldi and Pratelli (2005), imply that

$$\left(M_n, \tilde{V}_n, N_i^*\right)\Big| \, M_n, V_n \xrightarrow{w}_p (M, V, N_i) | \, M, V \tag{28}$$

as a convergence in probability of random measures on $(\mathscr{D}_{m\times m})^2 \times \mathscr{D}_m$, in the sense that

$$E\left[f\left(M_n, \tilde{V}_n, N_i^*\right)\Big| \, M_n, \tilde{V}_n\right] \xrightarrow{p} E\left[f\left(M, V, N_i\right) | \, M, V\right] \tag{29}$$

for any continuous bounded real f with matching domain.

The proof of the theorem is completed as in Theorems 5 and 6 of Hansen (2000), by using the following expansion which is uniform in $r \in [\underline{r}, \overline{r}]$: $F_{i,\lfloor nr\rfloor}^* = \tilde{F}_{in}(r) + o_{\mathbb{P}}(1)$ with

$$\tilde{F}_{in}(r) = \left\|(M_n(r) - M_n(r)M_n(1)^{-1}M_n(r))^{-1/2}(N_i^*(r) - M_n(r)M_n(1)^{-1}N_i^*(1))\right\|^2$$

and where $\mathbb{P}$ is the joint measure over the original and the bootstrap data. As $\tilde{F}_i(r)$ depends on the data only through $M_n, \tilde{V}_n$, it follows that

$$P^*(\max_{r\in[\underline{r},\overline{r}]} \tilde{F}_{in}(r) \leq \cdot) = P(\max_{r\in[\underline{r},\overline{r}]} \tilde{F}_{in}(r) \leq \cdot | M_n, \tilde{V}_n),$$

and since $\max_{r\in[\underline{r},\overline{r}]}\tilde{F}_{in}(r)|M_n,\tilde{V}_n \xrightarrow{w}_p \mathscr{F}_{i,\infty}|M,V$ by (28), with $\mathscr{F}_{i,\infty} := \sup_{r\in[\underline{r},\overline{r}]}\{\tilde{N}_i(r)'\tilde{M}(1)^{-1}\tilde{N}_i(r)\}$ $(i=1,2)$, also $\max_{r\in[\underline{r},\overline{r}]}\tilde{F}_{in}(r) \xrightarrow{w_*}_p \mathscr{F}_{i,\infty}|M,V$. Finally, as $\xrightarrow{\mathbb{P}}$ becomes $\xrightarrow{w^*}_p$ upon conditioning on the data, we conclude for $\mathscr{F}^*_{i,n} = \max_{r\in[\underline{r},\overline{r}]} F^*_{i,\lfloor nr\rfloor}$ that $\mathscr{F}^*_{i,n} \xrightarrow{w^*}_p \mathscr{F}_{i,\infty}|M,V$ on the special probability space. Then $\mathscr{F}^*_{i,n} \xrightarrow{w^*}_w \mathscr{F}_{i,\infty}|M,V$ in general. □

Proof of Theorem 4. Additionally to the notation introduced in the proof of Theorem 3, let $V_n := n^{-1}\sigma^{-2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x'_{nt}\varepsilon^2_{nt}$ and $X_n := \{x_{nt}\}_{t=1}^n$. Under Assumption $\mathcal{H}$, consider a common probability space where, for every $n\in\mathbb{N}$, the original and the bootstrap data is redefined such that (maintaining the notation),

$$\left(M_n, V_n, \tilde{V}_n, \tfrac{1}{n^{1/2}\sigma}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}\varepsilon_{nt}, \mathscr{F}_n\right) \xrightarrow{a.s.} (M,V,V,N,\mathscr{F}_\infty) \tag{30}$$

in $(\mathscr{D}_{m\times m})^3\times\mathscr{D}_m\times\mathbb{R}$, with $\mathscr{F}_\infty := \sup_{r\in[\underline{r},\overline{r}]}\{\tilde{N}(r)'\tilde{M}(r)^{-1}\tilde{N}(r)\}$ of eq. (22). On this space also $\mathscr{F}^*_{i,n} \xrightarrow{w^*}_p \mathscr{F}_{i,\infty}|M,V$ $(i=1,2)$ hold, by the proof of Theorem 3. Equivalently, $P^*(\mathscr{F}^*_{i,n}\le\cdot) \xrightarrow{p} P(\mathscr{F}_{i,\infty}\le\cdot|M,V)$ in $\mathscr{D}(\mathbb{R})$. Defining $Y_n := (M_n,\tilde{V}_n)$ and $Y_\infty := (M,V)$, we see that $(\mathscr{F}_n, Y_n, P(\mathscr{F}^*_{i,n}\le\cdot|D_n)) \xrightarrow{p} (\mathscr{F}_\infty, Y_\infty, P(\mathscr{F}_{i,\infty}\le\cdot|Y_\infty))$, $i=1,2$, on the special probability space. Whenever

$$P(\mathscr{F}_{i,\infty}\le\cdot|Y_\infty) = P(\grave{\mathscr{F}}_\infty\le\cdot|Y_\infty) \tag{31}$$

holds for some $i=1,2$, the convergence $(\mathscr{F}_n, Y_n, P^*(\mathscr{F}^*_{i,n}\le\cdot)) \xrightarrow{p} (\mathscr{F}_\infty, Y_\infty, P(\mathscr{F}_\infty\le\cdot|Y_\infty))$ on the special probability space implies that $(\mathscr{F}_n, Y_n, P^*(\mathscr{F}^*_{i,n}\le\cdot)) \xrightarrow{w} (\mathscr{F}_\infty, Y_\infty, P(\mathscr{F}_\infty\le\cdot|Y_\infty))$ on general probability spaces. As sample-path continuity of the conditional cdf $P(\grave{\mathscr{F}}_\infty\le\cdot|Y_\infty)$ is guaranteed by Proposition 3.2 of Linde (1989) applied conditionally on M,V, Theorem 2 becomes applicable. Specifically, for $i=1$, equality (31) holds if $M=V$, whereas for $i=2$ it holds independently of how M and V are related. Hence, in these cases the conclusions of Theorem 3 about asymptotic validity of the bootstrap procedures on average obtain from Theorem 2.

Let now Assumption $\mathcal{C}$ hold. Let the original and the bootstrap data be redefined on another probability space where (30) holds (and thus, $\mathscr{F}^*_{i,n} \xrightarrow{w^*}_p \mathscr{F}_{i,\infty}|M,V$, $i=1,2$), and additionally, the convergence in Assumption $\mathcal{C}$ holds as an a.s. convergence of random probability measures:

$$\left.\left(M_n, V_n, \tfrac{1}{n^{1/2}\sigma}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}\varepsilon_{nt}\right)\right| X_n \xrightarrow{w}_{a.s.} (M,V,N)\,|M,V.$$

The latter is possible by Corollary 5.12 of Kallenberg (1997), because the random measures in the previous display map measurably X_n and (M,V) to the Polish space of probability mesures on $(\mathscr{D}_{m\times m})^2\times\mathscr{D}_m$, equipped with the Prokhorov metric. Summarising, on the new probability space,

$$\begin{pmatrix} E\left[g\left(M_n, V_n, \tfrac{1}{n^{1/2}\sigma}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}\varepsilon_{nt}\right)\middle| X_n\right] \\ E\left[h\left(\mathscr{F}^*_{i,n}\right)\middle| D_n\right]\end{pmatrix} \xrightarrow{p} \begin{pmatrix} E\left[g(M,V,N)|\,M,V\right] \\ E\left[h(\mathscr{F}_{i,\infty})|\,M,V\right]\end{pmatrix}$$

for all continuous bounded real functions g, h. By expanding $F_{i,\lfloor nr\rfloor}$ similarly to $F^*_{i,\lfloor nr\rfloor}$ in the proof of Theorem 3 and choosing $g = k \circ s$, where s is the maximum over $r \in [\underline{r}, \overline{r}]$ of the leading term in that expansion and $k : \mathbb{R} \to \mathbb{R}$ is an arbitray continuous and bounded function, we can conclude that

$$\left(E\left(k\left(\mathscr{F}_n\right)\middle|\, X_n\right), E\left(h\left(\mathscr{F}^*_{i,n}\right)\middle|\, D_n\right)\right) \xrightarrow{p} \left(E\left(k\left(\mathscr{F}_\infty\right)\middle|\, M, V\right), E\left(h\left(\mathscr{F}_{i,\infty}\right)\middle|\, M, V\right)\right)$$

for all continuous bounded real functions k and h. As the distributions of the involved conditional expectations are fully determined by k, h and the distributions of $(\mathscr{F}_n, \mathscr{F}^*_{i,n}, X_n, D_n)$ and $(\mathscr{F}_\infty, \mathscr{F}_{i,\infty}, M, V)$, on general probability spaces the above convergence holds weakly. We notice that for $i = 1$, $\tilde{N}_1$ conditionally on M is independent of V (which can be removed from the conditioning in the second component of the limit).

As previously, sample-path continuity of the conditional distribution functions of $\mathscr{F}_\infty$ and $\mathscr{F}_{i,\infty}$ $(i = 1, 2)$, all conditional on M, V, is implied by Proposition 3.2 of Linde (1989) applied conditionally on M and V. This satisfies the continuity requirements of Corollary 1 (with $X_\infty := (M, V)$). The following cases emerge. Let $M = V$. Then the bootstrap based on $\mathscr{F}_n$ and $\mathscr{F}^*_{1,n}$ is asymptotically valid conditionally on X_n, by Corollary 1 with $\tau_\infty := \mathscr{F}_{1,\infty}$ and $X_\infty := M$. For both $M = V$ and $M \neq V$, the bootstrap based on $\mathscr{F}_n$ and $\mathscr{F}^*_{2,n}$ is asymptotically valid also conditionally on X_n, by Corollary 1 with $\tau_\infty := \mathscr{F}_{2,\infty}$ and $X_\infty := (M, V)$. □