1-1-2019

Bootstrapping the empirical distribution of a stationary process with change-point

Farid El Ktaibi
Zayed University

B. Gail Ivanoff
University of Ottawa, Canada

Bootstrapping the empirical distribution of a stationary process with change-point

Farid El Ktaibi

Department of Mathematics & Statistics
Zayed University
P.O. Box 144534
Abu Dhabi
United Arab Emirates
e-mail: Farid.Elktaibi@zu.ac.ae

and

B. Gail Ivanoff[*]

Department of Mathematics & Statistics
University of Ottawa
585 King Edward
Ottawa ON K1N 6N5
Canada
e-mail: givanoff@uottawa.ca

Abstract: When detecting a change-point in the marginal distribution of a stationary time series, bootstrap techniques are required to determine critical values for the tests when the pre-change distribution is unknown. In this paper, we propose a sequential moving block bootstrap and demonstrate its validity under a converging alternative. Furthermore, we demonstrate that power is still achieved by the bootstrap under a non-converging alternative. We follow the approach taken by Peligrad in [14], and avoid assumptions of mixing, association or near epoch dependence. These results are applied to a linear process and are shown to be valid under very mild conditions on the existence of any moment of the innovations and a corresponding condition of summability of the coefficients.

1. Introduction

Structural stability is typically a key component of time series models and corresponding inferential methods. Consequently, tests for structural change are crit-

ical when fitting any time series model to real-world data, and as a result there is an enormous literature on change-point detection. Somewhat surprisingly however, there has been relatively little attention paid to detecting a change in the marginal distribution of the time series. The most relevant references in this context are [7], [8] and most recently, [17]. The clear advantage of this approach is that the nature of the change (change in location, scale, covariance, etc.) need not be specified in the alternative, provided that it results in a change in the marginal distribution.

In the case of [7] and [8], the key to studying the asymptotic behaviour of the test statistics is a functional central limit theorem (FCLT) for the sequential empirical process both with and without a change point. A different approach is taken in [17] where more general forms of functional data are considered, with the empirical distribution considered as a special case. In [17], complicated tightness arguments are avoided by regarding the empirical distribution as a Hilbert space-valued functional, with weak convergence defined in terms of the L_2 norm of the Hilbert space. Appropriate test statistics are defined in [7], [8] and [17]. In [8] and [17], critical values for the test statistics are found via bootstrap techniques. In the case of [8], a weighted moving block bootstrap is shown to be valid under strong mixing conditions when the change evolves gradually over the observed period, while in [17], a disjoint block bootstrap (DBB) is proposed whose asymptotic behaviour is considered under a converging alternative, assuming L_1-near epoch dependence (NEP(1)) of the stationary sequence. In [7], there is little discussion of finding critical values, although the long memory linear process is discussed in some depth.

In this article, we take a closer look at the validity of the moving block bootstrap (MBB) in establishing critical values for detecting a change in the marginal distribution of a stationary time series using an approach due to Peligrad [14]. Theorem 2.2 of [14] establishes the validity of the MBB for the empirical distribution of a stationary process under straightforward moment conditions that do not involve any specific assumptions of mixing, association or near epoch dependence. Here we define a sequential version of the bootstrapped empirical process; our main results are a sequential version of Peligrad's Theorem 2.2 and an extension to a time series with a change point. As noted in [17], the most difficult aspect is establishing tightness of the bootstrapped empirical process in the function space $D(\mathbb{R} \times [0, 1])$.

There are several advantages of this approach. The MBB is known to be superior to the DBB of [17] in concrete applications (cf. [16], [11] and [12]). With Peligrad's moment conditions, we are able to establish almost sure weak convergence of the sequential bootstrapped empirical process on $D(\mathbb{R} \times [0, 1])$ under both converging and non-converging alternatives. Our main example, the causal linear process, is used to illustrate that the bootstrap is valid for processes that are not mixing or NEP(1). Consequently, we are able to apply the bootstrap to heavy-tailed models, which is of particular importance for examples in finance and economics. Further, by working with the empirical process we achieve more flexibility than is possible with the L_2 approach of [17] and are able to bootstrap any statistic that is a continuous functional of the sequential empirical process.

We demonstrate that Kolmogorov-Smirnov and Cramér-Von Mises test statistics achieve good power under both converging and non-converging alternatives.

We proceed as follows: we introduce the sequential moving block bootstrap in the next section and present a sequential version of Peligrad's CLT for the bootstrapped empirical process. In Section 3 we consider the behaviour of the bootstrapped empirical process when there is a change-point. It is shown in Theorem 3.2 that under a converging alternative, the bootstrapped sequential empirical process has the same asymptotic behaviour as the sequential empirical process without a change. On the other hand, Theorem 3.3 illustrates that a stronger normalization is needed when the alternative is not converging. In Section 4 we consider the asymptotic behaviour of Kolmogorov-Smirnov and Cramér-Von Mises test statistics under the null hypothesis and both converging and non-converging alternatives, and show that the bootstrap leads to consistent tests in all cases. Examples are given in Section 5, including an in-depth discussion of the linear model. In Section 6, simulations illustrate the performance of the tests in the case of both converging and non-converging alternatives. The tests will be seen to perform well even when first moments do not exist. Concluding comments and directions for further research are presented in Section 7, and all proofs appear in Section 8.

2. The bootstrap sequential empirical central limit theorem

As noted in the Introduction, our goal is to establish the validity of a sequential bootstrap technique that does not require conditions of mixing, association or near epoch dependence. To avoid these assumptions, we use the approach taken by Peligrad in [14], where sufficient conditions for the moving block bootstrap empirical CLT are expressed in terms of moments. In this section we review the moving block bootstrap and define a sequential bootstrapped empirical process. Our main result is Theorem 2.2, a sequential version of the bootstrap empirical CLT of [14].

We shall assume throughout that we have a strictly stationary ergodic stochastic process $(X_i, i \in \mathbf{Z})$ with marginal distribution F, defined on a probability space $(\Omega, \mathcal{F}, P)$. The empirical distribution function (edf) F_n is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}(X_i \le x), \ -\infty < x < \infty$$

and the *sequential* edf is defined as

$$F_{[ns]}(x) := \frac{1}{[ns]} \sum_{i=1}^{[ns]} \mathbf{I}(X_i \le x), \ -\infty < x < \infty, \ 0 \le s \le 1. \tag{1}$$

The empirical process is defined as

$$W_n(x) := \sqrt{n} \left(F_n(x) - F(x) \right)$$

and the sequential empirical process is

$$W_n(x,s) := \frac{[ns]}{\sqrt{n}} \left(F_{[ns]}(x) - F(x) \right)$$

Under appropriate regularity conditions, including some form of short memory (see, for example, [4], [5], [8]),

$$W_n(\cdot,\cdot) \xrightarrow{\mathbf{D}} W(\cdot,\cdot), \tag{2}$$

where $W(\cdot,\cdot)$ is a mean zero Gaussian process with covariance function

$$\sigma((x,s),(y,t)) = (s \wedge t) \sum_{i \in \mathbf{Z}} Cov\left(\mathbf{I}(X_0 \le x), \mathbf{I}(X_i \le y) \right),$$

and $\xrightarrow{\mathbf{D}}$ denotes weak convergence of random elements taking values in the space $D(\mathbb{R} \times [0,1])$ equipped with Skorokhod's J_1-topology (cf. [1] and [9] for more details).

We will be using the same version of the moving block bootstrap (MBB) as presented in [14] and [6]. Let l and k be positive integers and let $n = lk$. Now, extend a sample of size n by the first $l-1$ observations, namely, $X_1, \cdots, X_{l-1}$ to define the extended sequence X_{ni}, $i = 1, \cdots, n+l-1$ as follows:

$$X_{ni} := \left\{ \begin{array}{ll} X_i & \text{if} \quad 1 \le i \le n \\ X_{i-n} & \text{if} \quad n+1 \le i \le n+l-1. \end{array} \right.$$

Next, let $I_{n1}, I_{n2}, \cdots, I_{nk}$ be independent and identically distributed random variables each having uniform distribution on $\{1, 2, \cdots, n\}$. The intuitive idea behind the MBB is to concatenate k randomly chosen blocks of size l of the form $\{X_{nI_{nj}}, X_{n,I_{nj}+1}, \cdots, X_{n,I_{nj}+l-1}\}$, $1 \le j \le k$, and construct the bootstrap sample of size n,

$$\left(X_1^{(b)}, \cdots, X_n^{(b)} \right) = (X_{n,I_{n1}}, \cdots, X_{n,I_{n1}+l-1}, \cdots, X_{n,I_{nk}}, \cdots, X_{n,I_{nk}+l-1}). \tag{3}$$

The (non-sequential) bootstrapped empirical process is defined as follows: for $x \in \mathbb{R}$,

$$W_n^{(b)}(x) := \sqrt{n} \left[F_n^{(b)}(x) - F_n(x) \right], \tag{4}$$

where

$$F_n^{(b)}(x) := \frac{1}{n} \sum_{h=1}^{n} \mathbf{I}(X_i^{(b)} \le x) = \frac{1}{k} \sum_{j=1}^{k} \frac{1}{l} \sum_{i=I_{nj}}^{I_{nj}+l-1} \mathbf{I}(X_{ni} \le x) \tag{5}$$

is the bootstrapped empirical distribution.

Continuing with the sequential bootstrapped empirical process, define the ith block empirical distribution to be

$$F_{l,i}(x) := \frac{1}{l} \sum_{j=i}^{i+l-1} \mathbf{I}(X_{nj} \le x), \text{ and } F_n(x) = F_{n,1}(x).$$

Using the definition (5) of the bootstrapped empirical distribution and the fact that $\sum_{i=1}^{n} \mathbf{I}(I_{nj} = i) = 1$, when $n = lk$ we can rewrite the bootstrapped empirical process defined above in (4) as follows:

$$
\begin{aligned}
W_n^{(b)}(x) &= \sqrt{n}\left[F_n^{(b)}(x) - F_n(x) \right] \\
&= \sqrt{n}\left[\frac{1}{k}\sum_{j=1}^{k} F_{l,I_{nj}}(x) - F_n(x) \right] \\
&= \frac{\sqrt{n}}{k}\sum_{j=1}^{k}\left[\sum_{i=1}^{n} \mathbf{I}(I_{nj} = i)(F_{l,i}(x) - F_n(x)) \right].
\end{aligned}
\tag{6}
$$

The representation in (6) suggests the following definition for a sequential bootstrapped empirical process: for $(x, s) \in (\mathbb{R} \times [0, 1])$ and $n = lk$,

$$
\begin{aligned}
W_n^{(b)}(x, s) &= \frac{\sqrt{n}}{k}\sum_{j=1}^{[ks]}\left[\sum_{i=1}^{n} \mathbf{I}(I_{nj} = i)(F_{l,i}(x) - F_n(x)) \right] \\
&= \frac{l[ks]}{\sqrt{n}}\left(F_{l[ks]}^{(b)}(x) - F_n(x) \right),
\end{aligned}
\tag{7}
$$

where $F_{l[ks]}^{(b)}(x)$ is the bootstrapped empirical distribution based on $X_1^{(b)}, \ldots, X_{l[ks]}^{(b)}$.

As in [14], we assume the following relationship between the block lengths l_n and the number of blocks k_n, which allows the block size l_n to be arbitrarily close to $O(n^{1/3})$, observed by Künsch [11] to be the optimal length for the MBB. Following the notation used in [14], we write $a_n \ll b_n$ to indicate $a_n = O(b_n)$.

Assumption 2.1. Let (l_n) and (k_n) be sequences of natural numbers satisfying

$$
n^h \ll l_n \ll n^{\frac{1}{3}-a} \quad \text{for some} \quad 0 < h < \frac{1}{3} - a, \quad 0 < a < \frac{1}{3},
\tag{8}
$$

$l_n = l_{2^k}$ for $2^k \le n < 2^{k+1}$, $l_n \to \infty$ as $n \to \infty$ and $n = k_n l_n$.

Our first result is the following sequential version of Theorem 2.2 of [14].

Theorem 2.2. *Let $(X_n)_{n \in \mathbb{Z}}$ be a stationary sequence of random variables. Let l_n, k_n be sequences of natural numbers satisfying Assumption 2.1. Assume there are two constants C_1 and C_2 such that, for some $\gamma > 0$ and every $x < y$,*

$$
\sup_{n > m}\left| \sum_{i=m}^{n} Cov(\mathbf{I}(x < X_0 \le y), \mathbf{I}(x < X_i \le y)) \right| \le C_1 m^{-\gamma},
\tag{9}
$$

and for every $1 \le m \le n$,

$$
Var\left(\sum_{i=1}^{m} Y_{l_n i}^2(x, y) \right) \le C_2 m l_n^4,
\tag{10}
$$

where

$$Y_{l_n i}(x, y) := \sum_{j=i}^{i+l_n-1} \left(\mathbf{I}(x < X_j \le y) - (F(y) - F(x)) \right).$$

Then, defining $W_n^{(b)}(\cdot, \cdot)$ as in (7), as $n \to \infty$,

$$W_n^{(b)}(\cdot, \cdot) \xrightarrow{\mathbf{D}} W(\cdot, \cdot)$$

almost surely, where $W(\cdot, \cdot)$ is a Gaussian process with zero mean and covariance

$$Cov\left(W(x, s), W(y, t) \right) = (s \wedge t) \sum_{i \in \mathbb{Z}} Cov\left(\mathbf{I}(X_0 \le x), \mathbf{I}(X_i \le y) \right). \qquad (11)$$

The proof appears in Section 8.

Comments:

1. Assumption 2.1 requires that n be a multiple of l_n. In practice, if this is not the case then the number of blocks selected is $k_n = \left[\frac{n}{l_n} \right]$. It is onerous but straightforward to show that this does not affect the asymptotic behaviour of the resampled process. See [17] for example.
2. We note that that the sequential bootstrap is valid under the same conditions as in [14] for the usual empirical process. Although this is unsurprising, the proof of tightness is highly technical. Unfortunately, although the blocks are (conditionally) i.i.d., their definition changes with n. As a result, the straightforward argument used by van der Vaart and Wellner in [18] for sequential empirical processes based on a single sequence of i.i.d. random variables cannot be directly applied to the moving block bootstrap.
3. As observed by Peligrad in [14] and Radulovic in [16], the bootstrap process may converge in situations in which the original sequential empirical process does not, and vice versa. This point will be illustrated in the discussion in Section 5.

3. The bootstrap empirical central limit theorems with change-point

We now introduce the change-point model. Let $(Y_i, i \in \mathbb{Z})$ and $(Z_i, i \in \mathbb{Z})$ be strictly stationary ergodic sequences and let $\theta_n \in (0, 1]$. Let F and G be the distributions of Y_0 and Z_0, respectively. Borrowing the notation of [7], we write $\mathbf{X_n} := (X_1, ..., X_n) \in \mathbf{\Psi}_n(\theta_n, F, G)$ if

$$X_i = \begin{cases} Y_i & \text{for } 1 \le i \le [n\theta_n] \\ Z_i & \text{for } [n\theta_n] < i \le n. \end{cases} \qquad (12)$$

In the case of no change ($\theta_n = 1 \ \forall n$), we write $\mathbf{X_n} \in \mathbf{\Psi}_n(F)$.

Recall the sequential empirical distribution function defined in (1):

$$F_{[ns]}(x) = \frac{1}{[ns]} \sum_{i=1}^{[ns]} \mathbf{I}(X_i \le x).$$

Define, for $0 \le s \le 1$

$$H^{(n)}(x, s) := (s \wedge \theta_n)F(x) + (s - \theta_n)^+ G(x), \tag{13}$$

where $s^+ = \max(0, s)$. For $(x, s) \in \mathbb{R} \times [0, 1]$, let

$$
\begin{aligned}
W_n'(x, s) &:= \frac{1}{\sqrt{n}} \left([ns]F_{[ns]}(x) - nH^{(n)}(x, s) \right) \\
&= \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{[ns]} \mathbf{I}(X_i \le x) - nH^{(n)}(x, s) \right].
\end{aligned}
$$

We note $H^{(n)}(x, s) = sF(x)$ when there is no change ($\theta_n = 1$), and so under the null hypothesis $|W_n'(x, s) - W_n(x, s)| \le 1/\sqrt{n}$ and the processes are asymptotically equivalent.

Under appropriate regularity conditions on both the pre- and post-change stationary sequences (including some form of short memory) and assuming that $\theta_n \to \theta \in [0, 1]$, it may be shown that if $\mathbf{X_n} \in \mathbf{\Psi}_n(\theta_n, F, G)$, as $n \to \infty$, then

$$W_n'(\cdot, \cdot) \overset{\mathbf{D}}{\to} W^{(\theta)}(\cdot, \cdot), \tag{14}$$

where $W^{(\theta)}(\cdot, \cdot)$ is a centred Gaussian process with finite covariance

$$
\begin{aligned}
\sigma((x, s), (y, t)) &= (s \wedge t \wedge \theta) \sum_{i \in \mathbf{Z}} Cov\left(\mathbf{I}(Y_0 \le x), \mathbf{I}(Y_i \le x) \right) \\
&+ ((s \wedge t) - \theta)^+ \sum_{i \in \mathbf{Z}} Cov\left(\mathbf{I}(Z_0 \le x), \mathbf{I}(Z_i \le x) \right).
\end{aligned}
$$

See, for example, Theorem 2.4 of [5].

We now consider the behaviour of the sequential bootstrapped empirical process $W_n^{(b)}(\cdot, \cdot)$ defined in (7). We have two cases to consider, and we will use the following definition:

Definition 3.1. *We say that we have a* **converging alternative** *if $\theta_n \to 0$ or if $\theta_n \to 1$. The alternative is* **non-converging** *if $\theta_n \to \theta \in (0, 1)$.*

The so-called "converging alternative" means that the change takes place quite early or late in the observation period, and we shall see that consistency of the bootstrap depends on the rate at which θ_n converges to 0 or 1.

Theorem 3.2. *For the change-point model (12), assume that both stationary sequences $\{Y_i\}$ and $\{Z_i\}$ satisfy (9) and (10) and suppose that (l_n) and (k_n) are*

any sequences satisfying Assumption 2.1. Further, suppose that either $l_n\theta_n \to 0$ and $n\theta_n \to \infty$ or that $l_n(1 - \theta_n) \to 0$ and $n(1 - \theta_n) \to \infty$. Then, as $n \to \infty$,

$$W_n^{(b)}(\cdot, \cdot) \overset{\mathbf{D}}{\to} W'(\cdot, \cdot)$$

almost surely, where $W'(\cdot, \cdot)$ is a Gaussian process with zero mean and covariance

$$
\begin{aligned}
&Cov\left(W'(x, s), W'(y, t)\right) \\
&= \begin{cases}
(s \wedge t) \displaystyle\sum_{i \in \mathbb{Z}} Cov\left(\mathbf{I}(Z_0 \le x), \mathbf{I}(Z_i \le y)\right) & if \quad l_n\theta_n \to 0 \\[3mm]
(s \wedge t) \displaystyle\sum_{i \in \mathbb{Z}} Cov\left(\mathbf{I}(Y_0 \le x), \mathbf{I}(Y_i \le y)\right) & if \quad l_n(1 - \theta_n) \to 0.
\end{cases}
\end{aligned}
$$

From (14), we see that the limiting distribution is that of $W^{(1)}(\cdot, \cdot)$ when $\theta_n \to 1$, and $W^{(0)}(\cdot, \cdot)$ when $\theta_n \to 0$.

When the assumption of a converging alternative is not satisfied, the conditional covariances of the bootstrapped process diverge. In this case, we need a stronger normalization.

Theorem 3.3. *For the change-point model (12), assume that both stationary sequences $\{Y_i\}$ and $\{Z_i\}$ satisfy (9) and (10) and suppose that (l_n) and (k_n) are any sequences satisfying Assumption 2.1. Further, suppose that $\theta_n \to \theta \in (0, 1)$. Then, as $n \to \infty$,*

$$l_n^{-1/2} W_n^{(b)}(\cdot, \cdot) \overset{\mathbf{D}}{\to} L^{(\theta)}(\cdot, \cdot)$$

almost surely, where

$$L^{(\theta)}(x, s) = \sqrt{\theta(1 - \theta)}(F(x) - G(x))B(s), \tag{15}$$

and $B(\cdot)$ is a standard Brownian motion process on $[0, 1]$.

Remark 3.4. The significance of Theorem 3.3 is that while the bootstrapped test statistics defined in the next section diverge, they do so more slowly than the original test statistics, and so the bootstrap test still achieves power without a converging alternative. This will be made precise in Proposition 4.6 and Comment 4.7.

The proofs of Theorems 3.2 and 3.3 appear in Section 8.

4. Test statistics

Recalling the notation introduced in the preceding section, when there is an unknown change-point, the hypothesis and alternative are

$$
\begin{aligned}
H_0 &: \quad \{\exists F \quad \text{such that} \quad \mathbf{X}_n \in \mathbf{\Psi}_n(F)\} \\
H_1 &: \quad \{\exists \theta_n \in (0, 1), \ \exists F \ne G \ \text{such that} \ \mathbf{X}_n \in \mathbf{\Psi}_n(\theta_n, F, G)\}.
\end{aligned}
$$

In this case, the test statistics will be based on the following process:

$$V_n(x, s) := \frac{[ns](n - [ns])}{n^{\frac{3}{2}}} \left(F_{[ns]}(x) - F^*_{n-[ns]}(x) \right) \qquad (16)$$

where

$$F^*_{n-m}(x) = \frac{1}{n - m} \sum_{i=m+1}^{n} \mathbf{I}(X_i \leq x)$$

is the empirical distribution function based on $X_{m+1}, \ldots, X_n$. The process $V_n(\cdot, \cdot)$ compares the (suitably weighted) empirical distributions before and after $[ns]$, for $0 \leq s \leq 1$:

$$F_{[ns]}(x) - F^*_{n-[ns]}(x) = \frac{1}{[ns]} \sum_{i=1}^{[ns]} \mathbf{I}(X_i \leq x) - \frac{1}{n - [ns]} \sum_{i=[ns]+1}^{n} \mathbf{I}(X_i \leq x).$$

To test the pair (H_0, H_1), we use the following statistics:

- Weighted Kolmogorov-Smirnov statistic:

$$T_1 = \sup_{(x,s) \in \mathbb{R} \times [0,1]} |V_n(x, s)|.$$

- Weighted Cramér-Von Mises statistic:

$$T_2 = \int_0^1 \int_{\mathbb{R}} |V_n(x, s)|^2 \, dF_n(x) ds.$$

We reject the null hypothesis H_0 for large values of T_i, for $i = 1, 2$.

Proposition 4.1. *Under the null hypothesis H_0 and assuming that $W_n(\cdot, \cdot) \xrightarrow{\mathbf{D}} W(\cdot, \cdot)$, we have for every $c > 0$*

$$\lim_{n \to \infty} P\{T_1 > c\} = P \left\{ \sup_{(x,s) \in \mathbb{R} \times [0,1]} |W(x, s) - sW(x, 1)| > c \right\}, \qquad (17)$$

$$\lim_{n \to \infty} P\{T_2 > c\} = P \left\{ \int_0^1 \int_{\mathbb{R}} |W(x, s) - sW(x, 1)|^2 dF(x) ds > c \right\}, \qquad (18)$$

where $W(\cdot, \cdot)$ is the limiting process in (2).

The proof of Proposition 4.1 is identical to that of [5], Proposition 2.7.

We next deal with consistency of the test statistics T_1 and T_2. There are two cases of converging alternatives to consider: $\theta_n \xrightarrow{p} 0$ and $\theta_n \xrightarrow{p} 1$. The test statistics are consistent under a converging alternative provided that the rate of convergence of θ_n is slower than $1/\sqrt{n}$. The following is a slight sharpening of [5], Proposition 2.8. The proof appears in Section 8.

Proposition 4.2. *Suppose that $\mathbf{X}_n \in \mathbf{\Psi}_n(\theta_n, F, G)$ where $F \neq G$ and that $W'_n(\cdot, \cdot) \xrightarrow{\mathbf{D}} W^{(\theta)}(\cdot, \cdot).$*

(A) *Converging alternative:*

If the sequence $\{\theta_n : n \in \mathbb{N}\}$ satisfies either

1. $\theta_n \to 0$ and $\sqrt{n}\theta_n \to \infty$,

or

2. $\theta_n \to 1$ and $\sqrt{n}(1 - \theta_n) \to \infty$,

then $T_i \overset{P}{\to} \infty$ as $n \to \infty$, for $i = 1, 2$.

(B) *Non-converging alternative:*

If $\theta_n \to \theta \in (0, 1)$, *then*

1. $n^{-1/2}T_1 \overset{P}{\to} \theta(1 - \theta)\sup_{x \in \mathbb{R}}|F(x) - G(x)|$,

and

2. $n^{-1}T_2 \overset{P}{\to} \frac{(\theta(1-\theta))^2}{3}\int_{\mathbb{R}}(F(x) - G(x))^2 dH(x)$, *where* $H(x) = \theta F(x) + (1 - \theta)G(x)$.

4.1. The bootstrapped test statistics

Recall that the hypothesis and alternative are

$$H_0 \quad : \quad \{\exists F \quad \text{such that} \quad \mathbf{X}_n \in \mathbf{\Psi}_n(F)\}$$
$$H_1 \quad : \quad \{\exists \theta_n \in (0, 1), \ \exists F \neq G \text{ such that } \mathbf{X}_n \in \mathbf{\Psi}_n(\theta_n, F, G)\},$$

where $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\theta_n \in (0, 1)$ such that $\theta_n \to \theta \in [0, 1]$ as $n \to \infty$. Let $\{l_n\}$ and $\{k_n\}$ be sequences of natural numbers satisfying Assumption 2.1 such that $n = l_n k_n$ and suppose that $W_n'(\cdot, \cdot) \overset{\mathbf{D}}{\to} W^{(\theta)}(\cdot, \cdot)$. Suppressing dependence of l and k on n for notational convenience, recall the bootstrap sample of size n defined in (3):

$$\left(X_1^{(b)}, \ldots, X_n^{(b)}\right) = \left(X_{nI_{n1}}, \ldots, X_{n,I_{n1}+l-1}, \ldots, X_{nI_{nk}}, \ldots, X_{n,I_{nk}+l-1}\right),$$

where $I_{n1}, I_{n2}, \ldots, I_{nk}$ are independent and identically distributed random variables each having uniform distribution on $\{1, 2, \ldots, n\}$.

The testing procedure is based on $V_n^{(b)}(\cdot, \cdot)$, the sequential counterpart of the process $V_n(\cdot\cdot)$ introduced in (16), defined as follows:

$$V_n^{(b)}(x, s) = \frac{l[ks](k - [ks])}{k\sqrt{n}}\left(F_{l[ks]}^{(b)}(x) - F_{lk-l[ks]}^{(*,b)}(x)\right),$$

where $F_{l[ks]}^{(b)}(x)$ and $F_{lk-l[ks]}^{(*,b)}(x)$ are the bootstrapped empirical distributions based respectively on $X_1^{(b)}, \ldots, X_{l[ks]}^{(b)}$ and $X_{l[ks]+1}^{(b)}, \ldots, X_n^{(b)}$.

Recalling the definition of $W_n^{(b)}$ in (7), it is straightforward to show that

$$V_n^{(b)}(x, s) = W_n^{(b)}(x, s) - \frac{[ks]}{k}W_n^{(b)}(x, 1).$$

We define bootstrapped versions of the Kolmogorov-Smirnov and Cramér-Von Mises statistics as follows:

- Bootstrapped Kolmogorov-Smirnov statistic:

$$T_1^{(b)} = \sup_{(x,s)\in\mathbb{R}\times[0,1]} \left| V_n^{(b)}(x,s) \right|.$$

- Bootstrapped Cramér-Von Mises statistic:

$$T_2^{(b)} = \int_0^1 \int_{\mathbb{R}} \left| V_n^{(b)}(x,s) \right|^2 dF_n(x)ds.$$

The asymptotic behaviour of the test statistics follows from an application of the continuous mapping theorem to Theorems 2.2, Theorem 3.2, and Theorem 3.3.

First we deal with the null hypothesis.

Proposition 4.3. *Under the assumptions of Theorem 2.2*

$$T_1^{(b)} \xrightarrow{d} \sup_{(x,s)\in\mathbb{R}\times[0,1]} |W(x,s) - sW(x,1)|$$

$$T_2^{(b)} \xrightarrow{d} \int_0^1 \int_{\mathbb{R}} |W(x,s) - sW(x,1)|^2 \, dF(x)ds,$$

almost surely as $n \to \infty$, where $W(\cdot,\cdot)$ is the limiting process in Theorem 2.2.

Next we deal with a converging alternative ($\theta_n \to 0$ or 1).

Proposition 4.4. *Under the assumptions of Theorem 3.2*

$$T_1^{(b)} \xrightarrow{d} \sup_{(x,s)\in\mathbb{R}\times[0,1]} |W'(x,s) - sW'(x,1)|$$

$$T_2^{(b)} \xrightarrow{d} \int_0^1 \int_{\mathbb{R}} |W'(x,s) - sW'(x,1)|^2 \, d\tilde{H}(x)ds,$$

almost surely as $n \to \infty$, where $W'(\cdot,\cdot)$ is the limiting process in Theorem 3.2 and

$$\tilde{H} = \begin{cases} G & if \quad l_n\theta_n \to 0 \\ F & if \quad l_n(1-\theta_n) \to 0. \end{cases}$$

Comment 4.5. We remark that under the conditions of Proposition 4.1 and Proposition 4.3, the test statistics T_1 and T_2 and their bootstrapped counterparts, $T_1^{(b)}$ and $T_2^{(b)}$, converge respectively under H_0 to the same limits. On the other hand, provided that the conditions of Proposition 4.2(A) and Proposition 4.4 are satisfied, we note that under H_1 the test statistics based on the original variables diverge, whereas the bootstrapped statistics converge weakly to finite limits. This will allow us to tabulate critical values by constructing repeated moving block bootstrap samples. More precisely, the test will be consistent if for some constant $c > 0$,

$$\begin{cases} \theta_n \to 0 \\ \sqrt{n}\theta_n \to \infty \\ l_n\theta_n \ll n^{-c}, \end{cases} \quad \text{or} \quad \begin{cases} \theta_n \to 1 \\ \sqrt{n}(1-\theta_n) \to \infty \\ l_n(1-\theta_n) \ll n^{-c}, \end{cases} . \quad (19)$$

This will certainly be true, for instance, if $n^{h-1/2} \ll \theta_n \ll n^{-1/3}$ for some $0 < h < 1/6$ in the first case or analogously $n^{h-1/2} \ll 1 - \theta_n \ll n^{-1/3}$ in the second case.

Finally, we consider a non-converging alternative $(\theta_n \to \theta \in (0,1))$.

Proposition 4.6. *Under the assumptions of Theorem 3.3,*

$$l_n^{-1/2}T_1^{(b)} \quad \overset{d}{\to} \quad \sup_{(x,s)\in\mathbb{R}\times[0,1]} \left| L^{(\theta)}(x,s) - sL^{(\theta)}(x,1) \right|$$

$$\overset{d}{=} \quad \sqrt{\theta(1-\theta)} \left(\sup_{x\in\mathbb{R}} |F(x) - G(x)| \right) \left(\sup_{s\in[0,1]} |B^o(s)| \right)$$

$$l_n^{-1}T_2^{(b)} \quad \overset{d}{\to} \quad \int_0^1 \int_{\mathbb{R}} \left| L^{(\theta)}(x,s) - sL^{(\theta)}(x,1) \right|^2 dH(x)ds$$

$$\overset{d}{=} \quad \theta(1-\theta) \int_{\mathbb{R}} |F(x) - G(x)|^2 dH(x) \int_0^1 |B^o(s)|^2 ds,$$

almost surely as $n \to \infty$, where $L^{(\theta)}(\cdot,\cdot)$ is the limiting process in Theorem 3.3, $H(x) = \theta F(x) + (1-\theta)G(x)$, and B^o is a Brownian bridge on $[0,1]$.

Comment 4.7. Recalling that $l_n \ll n^{\frac{1}{3}-a}$ for some $a > 0$, if $\theta_n \to \theta \in (0,1)$ it is immediate from Proposition 4.2(B), that $l_n^{-1/2}T_1 \to \infty$ and $l_n^{-1}T_2 \to \infty$. Therefore, the bootstrapped statistics $T_1^{(b)}$ and $T_2^{(b)}$ diverge at a slower rate than the corresponding test statistics T_1 and T_2, ensuring that critical values tabulated from repeated bootstrap samples yield consistent tests. In fact, as will be seen in Section 6, simulations illustrate that good power is still achieved when the alternative is not converging.

5. Examples

For the test statistics T_1, T_2 and their bootstrapped counterparts to yield consistent tests of the hypothesis of no change, we need that

1. $W_n'(\cdot,\cdot) \overset{\mathbf{D}}{\to} W^{(\theta)}(\cdot,\cdot)$ (cf. (14));

and

2. The pre- and post-change sequences $\{Y_i\}$ and $\{Z_i\}$ both satisfy (9) and (10) of Theorem 2.2.

Note that the behaviour of the bootstrapped process $W_n^{(b)}$ depends only on item 2. above, and there are no additional assumptions needed on the relationship between the pre- and post-change sequences (cf. Theorems 3.2 and 3.3). However, for item 1. above, further information is required in order to ensure the asymptotic independence of the pre- and post-change increments of W_n' (for example, some form of joint short memory of $\{(Y_i, Z_i)\}$). This will be investigated in detail for the causal linear process, which was the principal motivation for

this work. First, we present two examples from [14] giving sufficient conditions for the validity of the bootstrap for mixing and associated sequences.

5.1. Mixing and associated sequences

Mixing sequences
Let $(X_0, X_1, X_2, ...)$ be stationary and strong mixing with mixing coefficients (α_n). Then conditions (9) and (10) of Theorem 2.2 are satisfied if for every $n \geq 1$,

$$\sum_{m>n} \alpha_m = O(n^{-\gamma})$$

for some $\gamma > 0$ ([14], Theorem 2.3).

Associated sequences
Let $(X_0, X_1, X_2, ...)$ be stationary and associated such that X_1 has a continuous bounded density. Then conditions (9) and (10) of Theorem 2.2 are satisfied if for every $n \geq 1$,

$$\sum_{m\geq n} [Cov(X_0, X_m)]^{1/3} = O(n^{-\gamma})$$

for some $\gamma > 0$ ([14], Theorem 2.4).

5.2. Causal linear processes

The causal linear process was the subject of both [6] and [5]. The validity of the MBB for the (non-sequential) empirical process under conditions similar to those of [3] was established in [6] and the linear process with a change-point was considered in [5]. Here, we combine these results to develop sufficient conditions for both items 1. and 2. above to hold, ensuring that the sequential MBB produces consistent tests for linear processes with change-point.

The stationary causal linear process is defined as follows: for $i \in \mathbb{Z}$,

$$X_i = \sum_{j\geq 0} a_j \xi_{i-j} \tag{20}$$

where $(\xi_j : j \in \mathbb{Z})$ is a stationary sequence of independent and identically distributed (i.i.d.) random variables and $(a_j : j \in \mathbb{N})$ is an absolutely summable sequence of constants. Summability of the coefficients ensures that if $E[\xi_0^2] < \infty$, the process has short memory in the sense that the covariances are summable, and more generally, if $E[|\xi_0|] < \infty$, the process is L_1-near epoch dependent (cf. [17]). However, as will be seen below, the linear model includes processes that are not mixing or NEP(1).

We begin by stating sufficient conditions for weak convergence of the empirical process $W'(\cdot, \cdot)$ both with and without a change (cf. (14)):

Assumptions 5.1.

1. Let $\{a_j, j \in \mathbb{Z}\}$ be a sequence of non-random weights, infinitely many of which are non-zero, satisfying

$$\sum_{j \geq 0} |a_j|^\gamma < \infty \quad \text{for some} \quad \gamma \in (0, 1].$$

2. There exist constants $C < \infty$ and $\Delta \in \left(\dfrac{2}{3}, 1\right]$ such that for all $u \in \mathbb{R}$

$$|E \exp(iu\xi_0)| \leq \frac{C}{(1 + |u|)^\Delta}.$$

3. $E[|\xi_0|^{4\gamma}] < \infty$, where $\gamma \in (0, 1]$ is as in 1 above.

Comments 5.2.

- Any linear process with Gaussian innovations and summable coefficients satisfies Assumptions 5.1, but such a process is not necessarily mixing, as illustrated by the following classic example due to Ibragimov: let the ξ_i's be i.i.d. $N(0, 1)$ and let a_i be the coefficient of z^i in the power series expansion of the function $h(z) = (1 - z)^p$, where $p > 4$ is non-integer. In this case, $|a_i| = O(i^{1-p})$ but (X_i) is not strong mixing.
- We emphasize that this model can include any sort of heavy-tailed innovations ξ_i, provided that $E[|\xi_0|^\delta] < \infty$ for some $\delta > 0$. Since it is implicit in the definition of L_1-near epoch dependence that the X_i's have a finite first moment (cf. [17]), the linear model above includes processes that are not NEP(1).
- Assumption 5.1.2 implies that the distribution function F_ξ of ξ_0 satisfies the Hölder condition $|F_\xi(x) - F_\xi(y)| < C|x - y|^\Delta$. It also implies that the distribution function of a partial sum of the $a_j\xi_{i-j}$ terms is differentiable with a bounded density satisfying a uniform Lipschitz condition, provided that sufficient terms with non-zero a_j are included in the moving average (cf. [3]). Obviously, the distribution function of X_0 is uniformly Lipschitz as well.
- The assumption that infinitely many coefficients (a_i) are non-zero is not required if F_ξ has a uniformly Lipschitz derivative. In this case, all the results that follow remain valid.

We now introduce the change-point model for the linear process. Let $\{\theta_n \in (0, 1], n \in \mathbb{N}\}$ be a convergent sequence with limit $\theta \in [0, 1]$. To define the causal linear process with a change-point at $[n\theta_n]$, consider the following stationary processes for $i \in \mathbb{Z}$ and $a_j^{(1)}, a_j^{(2)} \in \mathbb{R}$:

$$Y_i = \sum_{j \geq 0} a_j^{(1)} \xi_{i-j} \quad \text{and} \quad Z_i = \sum_{j \geq 0} a_j^{(2)} \xi'_{i-j},$$

where the vectors (ξ_i, ξ_i') are i.i.d. and both sequences $\{a_j^{(1)}\}$ and $\{a_j^{(2)}\}$ are absolutely summable. We do not make any assumption about the relation between ξ_i and ξ_i' – they can have any sort of dependence structure (and consequently, $\{Y_i\}$ and $\{Z_i\}$ need not be independent). Denote by F and G the respective distribution functions of Y_0 and Z_0.

The following result is Theorem 2.4 of [5]. Note that under the null hypothesis, $\theta_n = 1$ for all n.

Theorem 5.3. *Given the change-point model described above, assume that both linear processes $\{Y_i\}$ and $\{Z_i\}$ satisfy Assumptions 5.1. Further, assume that $\theta_n \to \theta$ where $\theta \in [0, 1]$. Then, for $\mathbf{X_n} \in \mathbf{\Psi}_n(\theta_n, F, G)$, as $n \to \infty$,*

$$W_n'(\cdot, \cdot) \overset{\mathbf{D}}{\to} W^{(\theta)}(\cdot, \cdot),$$

where $W^{(\theta)}(\cdot, \cdot)$ is defined in (14).

Next, we deal with weak convergence of the bootstrapped linear process $W_n^{(b)}$. However, we impose a slightly different set of assumptions than those for convergence of W_n'.

Assumptions 5.4.

1. Let $(a_j, j \in \mathbf{Z})$ be a sequence of non-random weights, infinitely many of which are non-zero, such that for some $\gamma \in (0, 1]$,

$$\sum_{j \geq 0} j|a_j|^\gamma = \sum_{j \geq 1} A_j(\gamma) < \infty,$$

where $A_j(\gamma) := \sum_{i \geq j} |a_i|^\gamma$.

2. There exist constants $C < \infty$ and $\Delta > 0$ such that for all $u \in \mathbb{R}$

$$|E \exp(iu\xi_0)| \leq \frac{C}{(1 + |u|)^\Delta}.$$

3. $E[|\xi_0|^{2\gamma}] < \infty$ where $\gamma \in (0, 1]$ is as in 1 above.

The following theorem follows from the proof of Theorem 2.5 in [6] and demonstrates the validity of the sequential moving block bootstrap.

Theorem 5.5. *If (X_n) is a stationary causal linear process satisfying Assumptions 5.4, then conditions (9) and (10) of Theorem 2.2 are satisfied. If, in addition, Assumption 2.1 is satisfied, then as $n \to \infty$,*

$$W_n^{(b)}(\cdot, \cdot) \overset{\mathbf{D}}{\to} W(\cdot, \cdot).$$

Corollary 5.6. *Given the change-point model described above, assume that both linear processes $\{Y_i\}$ and $\{Z_i\}$ satisfy Assumptions 5.4. Then both Theorems 3.2 and 3.3 obtain.*

The results of this subsection can be summarized as follows:

Corollary 5.7. *Given the change-point model described above for the causal linear process, assume that both linear processes $\{Y_i\}$ and $\{Z_i\}$ satisfy Assumptions 5.1 and 5.4, and that Assumption 2.1 holds. Then Propositions 4.3, 4.4 and 4.6 obtain and, in particular, for $\mathbf{X_n} \in \mathbf{\Psi}_n(\theta_n, F, G)$, the test statistics defined in Section 4 and their bootstrapped counterparts provide consistent tests of the hypothesis of no change.*

Comment 5.8. As noted in [6], it is not particularly surprising that Assumptions 5.1 and Assumptions 5.4 are not the same, since it is well known that a bootstrap CLT may be valid in situations when the original process does not satisfy a CLT, and vice versa. However, it is easily seen that both Assumptions 5.1 and Assumptions 5.4 are satisfied if Assumption 5.1.2 holds and $\{|a_j|\}_{j \in \mathbf{N}}$ is a non-increasing sequence such that $\sum_{j \geq 0} |a_j|^{\frac{\gamma}{2}} < \infty$ and $E[|\xi_0|^{2\gamma}] < \infty$ for some $\gamma \in (0, 1]$. As observed previously, these conditions include many linear sequences that are not mixing or NEP(1), and so cannot be handled by the results of [8] or [17].

6. Simulations

In this section, we illustrate the performance of the MBB tests in detecting changes under various scenarios. We begin by simulating the linear model to illustrate the performance of the tests proposed in Section 4 for both converging and non-converging alternatives. In addition, we consider both normal and Cauchy innovations to illustrate that the procedure performs well regardless of whether the innovations have a finite first moment.

To this end, we consider the following stationary autoregressive processes

$$\begin{cases} Y_i = \rho_1 Y_{i-1} + \xi_i \\ Z_i = \rho_2 Z_{i-1} + \xi'_i, \end{cases}$$

where the vectors of innovations (ξ_i, ξ'_i) are i.i.d. and $\rho_1, \rho_2 < 1$ ($a_j^{(i)} = \rho_i^j$, $i = 1, 2$). The change-point model satisfies

$$X_i = \begin{cases} Y_i & \text{if} & 1 \leq i \leq [n\theta_n] \\ Z_i & \text{if} & [n\theta_n] < i \leq n, \end{cases}$$

where $\theta_n \to \theta \in [0, 1]$.

In our first two examples we consider normal and Cauchy innovations. In both cases, we investigate separately changes in location or scale of the innovations under converging ($\theta_n = 0.08$) and non-converging alternatives ($\theta_n = 0.5$), illustrating the performance of both test statistics, the Kolmogorov-Smirnov (K.S) and Cramér-Von Mises (C.V.M). In all our examples, the tests are carried out at a nominal level of significance $\alpha = 5\%$ with 500 bootstrap replications used to determine the appropriate critical value in each case. Each simulation was repeated 400 times for the analysis of the power.

It was found that the converging alternative was much more sensitive to the relation between n and l than the non-converging alternative. In the case of a

converging alternative ($\theta_n = 0.08$), the sample size used was $n = 15,000$ with $l = 15$ and $k = 1,000$; for the non-converging alternative ($\theta_n = 0.5$), the sample size was $n = 10,000$ with $l = 25$ and $k = 400$.

6.1. Example 1

Here we investigate the performance of our test statistics in detecting a change in an AR(1) process with normal innovations. We consider changes in the mean and the variance of the innovations.

• Change in the mean of the innovations:

In this case, we consider the following model with $\rho_1 = \rho_2 = 0.5$:

$$\begin{cases} Y_i = 0.5Y_{i-1} + \xi_i \\ \\ Z_i = 0.5Z_{i-1} + \xi_i' \end{cases} \quad \text{where } \xi_i \sim \mathcal{N}(0,1) \text{ and } \xi_i' \sim \mathcal{N}(\mu, 1).$$

We consider $\mu = 0$ under the null hypothesis and $\mu = -0.5, -0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4, 0.5$ under the alternative.

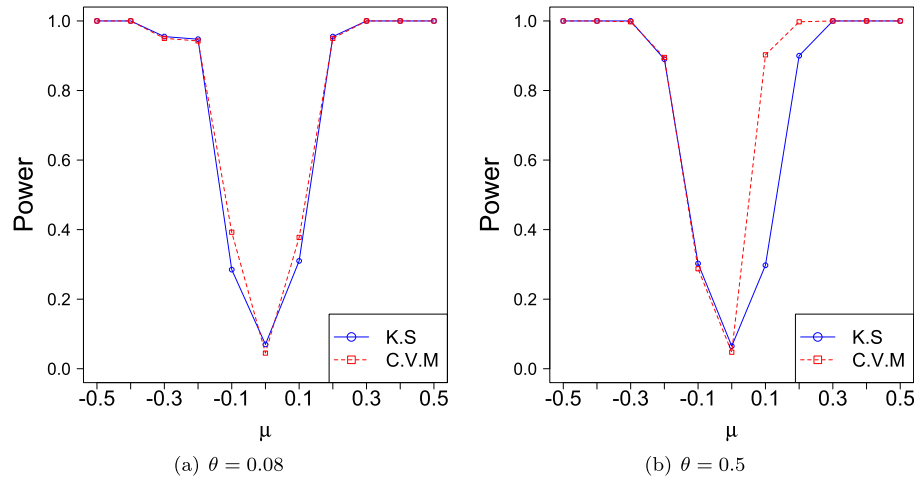The performance of the K.S and C.V.M test statistics for this case are illustrated in Fig. 1.

(a) $\theta = 0.08$ (b) $\theta = 0.5$

FIG 1. *Detection of a change in the mean of the normal innovations*

• Change in the variance of the innovations:

In this case, we consider the following model with $\rho_1 = \rho_2 = 0.5$:

$$\begin{cases} Y_i = 0.5Y_{i-1} + \xi_i \\ \\ Z_i = 0.5Z_{i-1} + \xi_i' \end{cases} \quad \text{where } \xi_i \sim \mathcal{N}(0,1) \text{ and } \xi_i' \sim \mathcal{N}(0, \sigma^2).$$

We consider $\sigma = 1$ under the null hypothesis and $\sigma = 0.5, 0.6, 0.7, 0.8, 0.9, 1.1,$ $1.2, 1.3, 1.4, 1.5$ under the alternatives. The empirical size and the power performance of the tests are illustrated in Fig. 2.
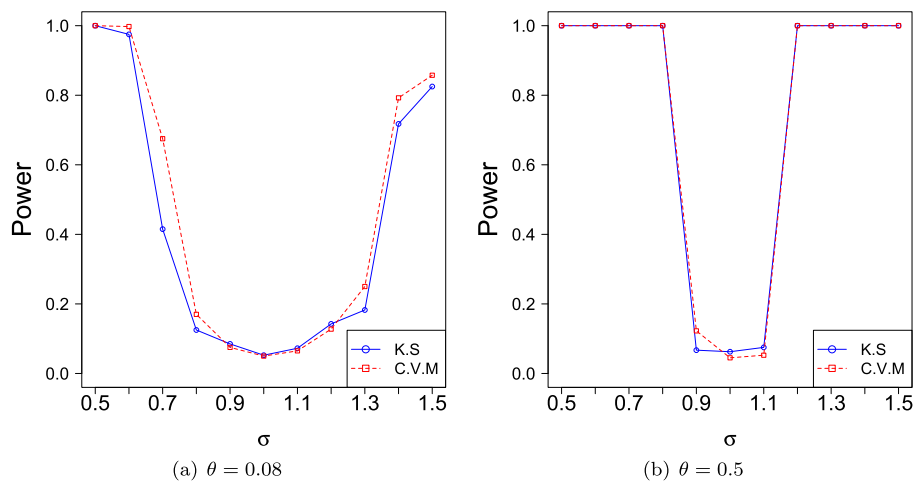

(a) $\theta = 0.08$ (b) $\theta = 0.5$

Fig 2. *Detection of a change in the variance of the normal innovations*

6.2. Example 2

Using the same parameters as in the preceding example, we will be considering now a stationary autoregressive model with Cauchy innovations.

• Change in the location parameter:

The model to be considered in this case is as follows:

$$\begin{cases} Y_i = 0.5Y_{i-1} + \xi_i \\ \\ Z_i = 0.5Z_{i-1} + \xi_i' \end{cases} \quad \text{where } \xi_i \sim \mathcal{C}(0,1) \text{ and } \xi_i' \sim \mathcal{C}(a,1).$$

We consider here the location parameter $a = 0$ under the null hypothesis. On the other hand, a will take the following values under the alternatives $a = -0.5, -0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4, 0.5$. The empirical size and the power performance of the K.S. and C.V.M. statistics are shown in Fig. 3.
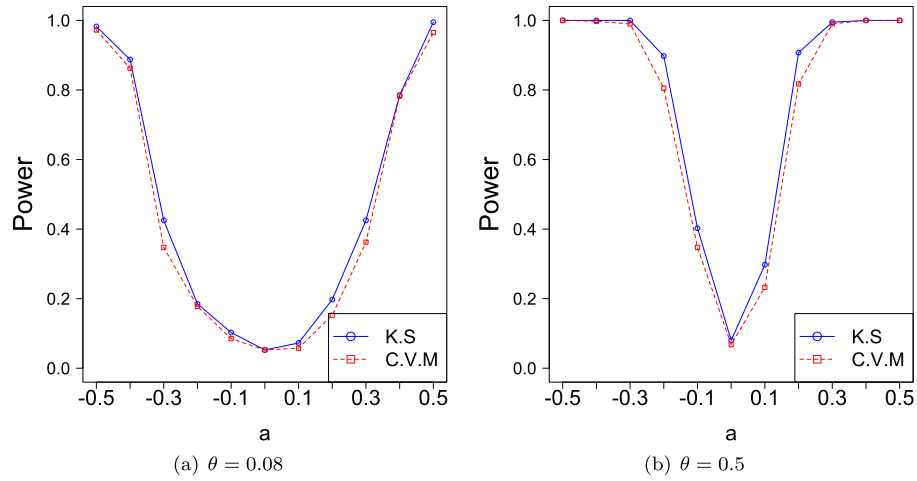
(a) $\theta = 0.08$ (b) $\theta = 0.5$

FIG 3. *Detection of a change in the location of the Cauchy innovations*

- Change in the scale parameter:

In this case, we consider the following model:

$$
\begin{cases}
Y_i = 0.5Y_{i-1} + \xi_i \\
Z_i = 0.5Z_{i-1} + \xi_i',
\end{cases}
\quad \text{where } \xi_i \sim \mathcal{C}(0,1) \text{ and } \xi_i' \sim \mathcal{C}(0,b).
$$

Fig. 4 deals with a change in the scale of the Cauchy innovations of the autoregressive model. In this case, we assume that the scale parameter is $b = 1$ under the null hypothesis and $b = 0.5, 0.6, 0.7, 0.8, 0.9, 1.1, 1.2, 1.3, 1.4, 1.5$ under the alternatives.
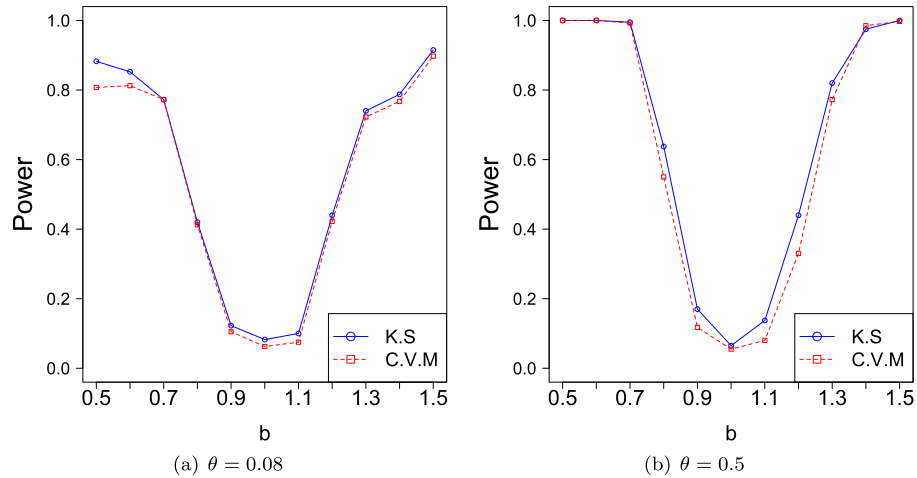

(a) $\theta = 0.08$ (b) $\theta = 0.5$

FIG 4. *Detection of a change in the scale of the Cauchy innovations*

Examples 1 and 2 illustrate that our unified approach allows us to easily detect changes in location or scale in the marginal distribution of a linear model. In all cases, we can see that the rejection rate under the null hypothesis is close to the nominal level of significance $\alpha = 0.05$ and that we achieve good power under the alternatives. We note that contrary to what is frequently observed, the Cramér-Von Mises statistic does not consistently outperform the Kolmogorov-Smirnov statistic. In both cases, normal and Cauchy innovations, we remark a notable improvement of the performance of the test under the non-converging alternative, despite the fact that the bootstrap test statistics diverge.

The large sample sizes used in Examples 1 and 2 are consistent with financial data. Next we consider smaller sample sizes and various block lengths.

6.3. Example 3

Analogously to the first example, we examine the performance of the testing procedure for smaller sample sizes n and various values of l when detecting a change in the mean of the normal innovations from $\mu_0 = 0$ to $\mu_1 = 0.5$ with $\theta_n = 0.5$. The empirical power and size are displayed in Table 1.

We observe that the tests still perform very well with much smaller sizes under the non-converging alternative. We can see that for all combinations of l and k, the power achieved is very good. However, the empirical size is higher than the nominal size for highly correlated processes ($\rho = 0.9$) and small values of l. This is an unsurprising issue of the bootstrapped tests since the short blocks do not accurately reflect the stronger dependence structure and consequently the critical values obtained by the bootstrap techniques are too small.

6.4. Example 4

In this example, we will present simulations that illustrate the effect of a change in the tail behaviour of the marginal distribution. To this end, we consider the following model:

$$\begin{cases} Y_i = \rho Y_{i-1} + \xi_i \\ Z_i = \rho Z_{i-1} + \xi_i', \end{cases} \quad \text{where } \xi_i \sim \mathcal{N}(0,1) \text{ and } \xi_i' \sim \mathcal{C}(0,1).$$

As in the previous example, $\theta_n = 0.5$ and we consider various values of n and l. Table 2 shows that the test performs very well in this case. However, the empirical power decreases as the length of blocks increases, since as observed in the preceding example, the bootstrapped critical values are too small with shorter blocks when the observations are highly correlated.

TABLE 1
Performance of K.S and C.V.M in small sample sizes

		Empirical size for K.S and C.V.M		
		$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
	$l = 1, k = 300$	0.086/0.078	0.435/0.389	0.993/0.998
	$l = 2, k = 100$	0.067/0.054	0.265/0.224	0.932/0.933
$n = 200$	$l = 4, k = 50$	0.078/0.059	0.159/0.130	0.759/0.704
	$l = 5, k = 40$	0.068/0.050	0.134/0.108	0.668/0.623
	$l = 8, k = 25$	0.064/0.052	0.096/0.073	0.461/0.414
	$l = 10, k = 20$	0.080/0.052	0.092/0.083	0.376/0.339
	$l = 1, k = 300$	0.071/0.082	0.412/0.352	0.999/0.998
	$l = 2, k = 150$	0.068/0.065	0.259/0.211	0.961/0.942
	$l = 3, k = 100$	0.079/0.058	0.189/0.165	0.851/0.803
$n = 300$	$l = 4, k = 75$	0.073/0.047	0.111/0.112	0.722/0.682
	$l = 5, k = 60$	0.047/0.033	0.129/0.109	0.666/0.616
	$l = 10, k = 30$	0.064/0.040	0.101/0.073	0.386/0.337
	$l = 1, k = 400$	0.090/0.080	0.419/0.385	0.997/1.000
	$l = 2, k = 200$	0.066/0.065	0.247/0.214	0.938/0.937
	$l = 4, k = 100$	0.054/0.059	0.145/0.108	0.760/0.706
$n = 400$	$l = 5, k = 80$	0.058/0.052	0.111/0.103	0.666/0.590
	$l = 8, k = 50$	0.069/0.046	0.097/0.080	0.474/0.420
	$l = 10, k = 40$	0.084/0.058	0.086/0.073	0.372/0.316
	$l = 16, k = 25$	0.068/0.049	0.078/0.064	0.257/0.227
	$l = 1, k = 500$	0.102/0.087	0.425/0.371	0.999/0.997
	$l = 2, k = 250$	0.069/0.047	0.256/0.224	0.950/0.937
$n = 500$	$l = 4, k = 125$	0.067/0.061	0.153/0.138	0.758/0.699
	$l = 5, k = 100$	0.076/0.056	0.115/0.090	0.710/0.613
	$l = 10, k = 50$	0.062/0.042	0.072/0.067	0.385/0.343
	Empirical power for K.S and C.V.M when detecting a change in the mean			
		$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
	$l = 1, k = 300$	0.9400/0.9775	0.9975/1	1/1
	$l = 2, k = 100$	0.9250/0.9550	0.9975/0.9975	1/1
$n = 200$	$l = 4, k = 50$	0.9250/0.9550	0.9925/0.9950	1/1
	$l = 5, k = 40$	0.9450/0.9275	0.9850/0.9875	1/1
	$l = 8, k = 25$	0.8925/0.9050	0.9700/0.9650	1/1
	$l = 10, k = 20$	0.9000/0.9275	0.9675/0.9700	1/1
	$l = 1, k = 300$	0.9925/0.9975	1/1	1/1
	$l = 2, k = 150$	0.9900/0.9950	1/1	1/1
$n = 300$	$l = 3, k = 100$	0.9825/0.9875	1/1	1/1
	$l = 4, k = 75$	0.9975/0.9950	1/1	1/1
	$l = 5, k = 60$	0.9875/0.9925	1/1	1/1
	$l = 10, k = 30$	0.9825/0.9925	0.9925/0.995	1/1
	$l = 1, k = 400$	0.9975/1	1/1	1/1
	$l = 2, k = 200$	0.9950/0.9950	1/1	1/1
	$l = 4, k = 100$	0.9975/0.9975	1/1	1/1
$n = 400$	$l = 5, k = 80$	0.9975/0.9975	1/1	1/1
	$l = 8, k = 50$	1/0.9975	1/0.9975	1/1
	$l = 10, k = 40$	1/1	1/1	1/1
	$l = 16, k = 25$	0.9900/1	1/1	1/1
	$l = 1, k = 500$	1/1	1/1	1/1
	$l = 2, k = 250$	1/1	1/1	1/1
$n = 500$	$l = 4, k = 125$	1/1	1/1	1/1
	$l = 5, k = 100$	1/1	1/1	1/1
	$l = 10, k = 50$	1/1	1/1	1/1

TABLE 2
Empirical power for K.S and C.V.M when detecting a change from Normal to Cauchy innovations

		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.9$
$n = 200$	$l = 1, k = 300$	0.6750/0.7775	0.9775/0.9925	1/1
	$l = 2, k = 100$	0.6375/0.6750	0.8825/0.9050	1/1
	$l = 4, k = 50$	0.4975/0.5025	0.7625/0.7375	0.9800/0.9900
	$l = 5, k = 40$	0.4825/0.4550	0.7125/0.7075	0.9800/0.9825
	$l = 8, k = 25$	0.4625/0.4525	0.6225/0.5450	0.9200/0.8825
	$l = 10, k = 20$	0.4950/0.4775	0.5500/0.5000	0.8425/0.7900
$n = 300$	$l = 1, k = 300$	0.8925/0.9325	0.9950/1.0000	1/1
	$l = 2, k = 150$	0.8500/0.9200	0.9850/0.9925	1/1
	$l = 3, k = 100$	0.7800/0.8425	0.9500/0.9625	1/1
	$l = 4, k = 75$	0.7100/0.7425	0.9275/0.9325	1/0.9975
	$l = 5, k = 60$	0.7025/0.7600	0.9100/0.9125	0.9950/0.9975
	$l = 10, k = 30$	0.7225/0.7200	0.8100/0.7500	0.9500/0.9425
$n = 400$	$l = 1, k = 400$	0.9650/0.9875	0.9975/1	1/1
	$l = 2, k = 200$	0.9525/0.9725	0.9950/0.9950	1/1
	$l = 4, k = 100$	0.9050/0.9225	0.9825/0.9850	1/1
	$l = 5, k = 80$	0.9050/0.9200	0.9750/0.9700	1/1
	$l = 8, k = 50$	0.8725/0.8650	0.9425/0.9575	0.9975/0.9975
	$l = 10, k = 40$	0.8775/0.8900	0.9300/0.9175	0.9875/0.9800
	$l = 16, k = 25$	0.8475/0.8800	0.9050/0.9025	0.9350/0.9050
$n = 500$	$l = 1, k = 500$	0.9975/0.9975	1/1	1/1
	$l = 2, k = 250$	0.9825/0.9950	1/1	1/1
	$l = 4, k = 125$	0.9825/0.9875	0.9925/0.9975	1/1
	$l = 5, k = 100$	0.9800/0.9825	0.9925/0.9950	1/1
	$l = 10, k = 50$	0.9625/0.9750	0.9850/0.9975	0.9950/0.9925

6.5. *Example 5*

Since the applicability of the results presented in this paper go beyond the scope of linear processes, one can analyse the performance of our testing approach in the nonlinear case. This example is of particular interest, since both the pre- and post-change distributions are defined to have mean 0 and variance 1. First we introduce a simple ARCH process defined recursively by $Z_i = \sigma_i \xi_i'$ where $\sigma_i^2 = (1 - a) + a Z_{i-1}^2$ with $0 \leq a < 1$. Figure 5 displays a change from an i.i.d sequence to a simple ARCH process:

$$\begin{cases} Y_i = \xi_i \\ Z_i = \sigma_i \xi_i', \end{cases} \quad \text{where } \xi_i \sim \mathcal{N}(0, 1) \text{ and } \xi_i' \sim \mathcal{N}(0, 1).$$

Figure 6 illustrates the following change:

$$\begin{cases} Y_i = \sigma_i^{(1)} \xi_i \\ Z_i = \sigma_i^{(2)} \xi_i', \end{cases} \quad \text{where } \xi_i \sim \mathcal{N}(0, 1) \text{ and } \xi_i' \sim \mathcal{N}(0, 1).$$

$\sigma_i^{(1)}$ and $\sigma_i^{(2)}$ are consecutively obtained by taking $a = 0.5$ in the first case and letting $a = 0.1, \cdots, 0.9$ for the second.
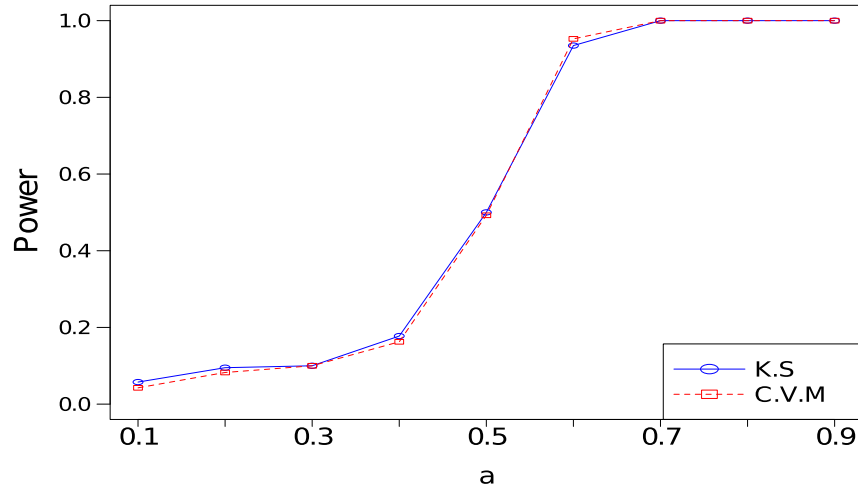
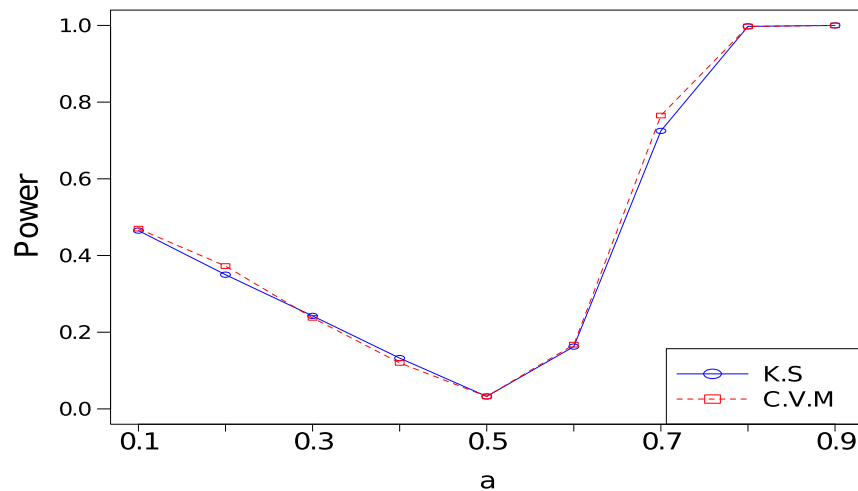Fɪɢ 5. *Detection of a change from an independent sequence to an ARCH process*

Fɪɢ 6. *Detection of a change in ARCH processes*

As in the previous examples, the simulations were made at a nominal level $\alpha = 5\%$ with 500 bootstrap replications, and each simulation was performed 400 times to get the empirical power under a non-converging alternative ($\theta_n = 0.5$). The sample size used in this case was $n = 5000$ with $l = 10$ and $k = 500$.

Figures 5 and 6 show that our procedure can be used to detect changes from an i.i.d sequence to an ARCH process or from an ARCH process to another one in case of moderate sample sizes. Although the usual distribution-free tests are applicable for i.i.d. random variables, our results show that the bootstrap works well even if the assumption that the pre-change process is i.i.d. has not

been made. In this case, the power increases as the coefficient a increases. This makes sense, since larger values of a cause more dependence in the post-change process, and this, in turn, is reflected in the post-change marginal distribution. The second change model shows that the empirical size is very close to the nominal level and the test statistics achieve an acceptable power for small values of a. It also shows much better performance for large values of a. In both cases, the performance of the test statistics (Kolmogorov-Smirnov, Cramér-Von Mises) is virtually identical.

7. Conclusion

In this article, we have defined a sequential version of the moving block bootstrap and demonstrated its validity in detecting a change-point in a stationary sequence under simple moment conditions. These results have been applied to a linear model to illustrate that the MBB can be applied to processes that are not mixing, associated, or near epoch dependent. Simulations illustrate the performance of the procedure in detecting changes in both linear and non-linear models.

There are many open questions beyond the scope of this paper that are of interest for further research:

- In fact, Theorem 2.4 of [5] is slightly more general than as stated here in Theorem 5.3. In the case of the linear process, it is possible to prove that (14) holds for random values of θ_n, provided that $\theta_n \to_P \theta$, where $\theta \in [0, 1]$ is fixed. The random change-point $[n\theta_n]$ can be either independent or data-dependent (for details, see [5]). It would be useful to investigate conditions under which the MBB can be extended to a random change-point.
- Estimation of the value of θ_n should be considered.
- The results obtained here should be extended to processes with multiple change points.
- There are many interesting models of stationary random fields that generalize both mixing and linear sequences (cf. [10]). The block bootstrap as well has been extended to planar processes. The methods developed here could be generalized to detect a planar change-point or change-set. This question is currently under investigation.

8. Proofs

8.1. Proof of Theorem 2.2

In what follows, for notational convenience we write $l = l_n$ and $k = k_n$ and fix a realization of the stochastic process, $\{x_i\}$. Hence, the resampling mechanism becomes the unique source of randomness. We also consider the triangular array $\{x_{ni}\}$ defined as in Section 2:

$$x_{ni} := \begin{cases} x_i & \text{if} \quad 1 \leq i \leq n \\ x_{i-n} & \text{if} \quad n+1 \leq i \leq n+l-1. \end{cases}$$

Some definitions will be needed in the sequel.

The ith block sample mean:

$$\overline{x}_{li} = \frac{1}{l} \sum_{j=i}^{i+l-1} x_{nj},$$

and the sample mean:

$$\overline{x}_n = \frac{1}{n} \sum_{j=1}^{n} x_j = \frac{1}{n} \sum_{i=1}^{n} \overline{x}_{li}.$$

The bootstrapped sample mean is then defined as

$$\overline{x}_n^{(b)} = \frac{1}{n} \sum_{i=1}^{n} x_i^{(b)},$$

where $(x_1^{(b)}, \ldots, x_n^{(b)}) = (x_{nI_{n1}}, \ldots, x_{n,I_{n1}+l-1}, \ldots, x_{nI_{nk}}, \ldots, x_{n,I_{nk}+l-1})$ is defined as in Section 2.

We shall also use the notation:

$$f_{li}(x) = \frac{1}{l} \sum_{j=i}^{i+l-1} \mathbf{I}(x_{nj} \leq x), \qquad f_n(x) = f_{n1}(x)$$

$$f_{li}(x,y) = f_{li}(x) - f_{li}(y), \qquad f_n(x,y) = f_{n1}(x,y).$$

Define now, for $x \in \mathbb{R}$, $Z_n^{(b)}(x)$ to be a sample-based version of the bootstrapped empirical process by replacing X_{ni} by x_{ni} in (6). Hence,

$$
\begin{aligned}
Z_n^{(b)}(x) &= \sqrt{n}(f_n^{(b)}(x) - f_n(x)) \\
&= \frac{\sqrt{n}}{k} \sum_{j=1}^{k} \left[\sum_{i=1}^{n} \mathbf{I}(I_{nj} = i)(f_{li}(x) - f_n(x)) \right],
\end{aligned}
\tag{21}
$$

where $f_n^{(b)}(x)$ is defined by replacing X_{ni} by x_{ni} in (5).

We also define a sample-based version of the sequential bootstrapped empirical process, by replacing X_{ni} by x_{ni} in (7), for $(x,s) \in (\mathbb{R} \times [0,1])$ as

$$
Z_n^{(b)}(x,s) = \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \left[\sum_{i=1}^{n} \mathbf{I}(I_{nj} = i)(f_{li}(x) - f_n(x)) \right].
\tag{22}
$$

We should note again that the terms in the partial sum are independent and identically distributed.

To prove Theorem 2.2, we proceed with a sequence of propositions that are sequential versions of the results in Section 3 of [14]. Recall that we use C to denote a generic constant that may change at each appearance.

Our first proposition is a straightforward generalization of Proposition 3.1 of [14] and so is presented without proof.

Proposition 8.1. *Let $\{x_i, i \geq 1\}$ be a bounded sequence of real numbers. Let k and l be integers such that $n = kl$ and*

$$\frac{l^2}{n} \to 0 \quad as \quad n \to \infty. \tag{23}$$

For each n, let $\{I_{n1}, I_{n2}, \ldots, I_{nk}\}$ be i.i.d uniform on $\{1, 2, \ldots, n\}$ and assume that

$$V_n = \frac{1}{k} \sum_{i=1}^{n} \left(\overline{x}_{li} - \overline{x}_n\right)^2 \to \sigma^2 > 0 \quad as \quad n \to \infty. \tag{24}$$

Then, for $0 \leq s < t \leq 1$

$$\frac{\sqrt{n}}{k} \sum_{j=[ks]+1}^{[kt]} \sum_{i=1}^{n} \mathbf{I}(I_{nj} = i) \left(\overline{x}_{li} - \overline{x}_n\right) \xrightarrow{d} \mathcal{N}(0, (t-s)\sigma^2) \quad as \quad n \to \infty. \tag{25}$$

The following proposition proves the convergence of finite dimensional distributions $Z_n^{(b)}(\cdot, \cdot)$. We shall first define for x, y in $\mathbb{R}$

$$V_n(x, y) = \frac{1}{k} \sum_{i=1}^{n} (f_{li}(x, y) - f_n(x, y))^2. \tag{26}$$

Proposition 8.2. *Let $\{x_i, 1 \leq i \leq n\}$ and $\{I_{ni}, 1 \leq i \leq n\}$ be as in Proposition 8.1, and assume that $0 \leq x_i \leq 1$ for every $0 \leq i \leq n$. Let k and l be integers such that $n = kl$ and $l^2/n \to 0$ as $n \to \infty$. Assume also for every $x, y \in [0, 1]$ that*

$$\lim_{n \to \infty} V_n(x, y) \quad exists. \tag{27}$$

Then, for every x, y in $[0, 1]$

$$\lim_{n \to \infty} \frac{1}{k} \sum_{i=1}^{n} (f_{li}(x) - f_n(x))(f_{li}(y) - f_n(y)) = \sigma(x, y) \tag{28}$$

exists and for every $(z_1, \ldots, z_p, s_1, \ldots, s_p) \in [0, 1]^{2p}$

$$\left(Z_n^{(b)}(z_1, s_1), \ldots, Z_n^{(b)}(z_p, s_p)\right) \xrightarrow{d} \left(\mathcal{N}_{z_1, s_1}, \ldots, \mathcal{N}_{z_p, s_p}\right) \quad as \quad n \to \infty, \tag{29}$$

where $\{\mathcal{N}_{z_i, s_i}\}_{1 \leq i \leq p}$ are zero-mean Gaussian with covariance function

$$\sigma_{ij} = (s_i \wedge s_j)\sigma(z_i, z_j).$$

Proof. Recalling the definition of $Z_n^{(b)}(x)$ in (21), exactly as in the proof of Proposition 3.3 of [14] we have that the limit in (28) exists and

$$Cov\left(Z_n^{(b)}(x), Z_n^{(b)}(y)\right) = \frac{1}{k} \sum_{i=1}^{n} (f_{li}(x) - f_n(x))(f_{li}(y) - f_n(y)). \tag{30}$$

Using now the representation in (22), we can see that for $0 \leq s < t \leq 1$

$$Cov\left(Z_n^{(b)}(x,s), Z_n^{(b)}(y,t)\right) \quad = \quad \frac{[ks]}{k^2}\sum_{i=1}^{n}(f_{li}(x) - f_n(x))(f_{li}(y) - f_n(y))$$

$$\rightarrow \quad s\sigma(x,y) \quad \text{as} \quad n \rightarrow \infty.$$

Suppose now, for instance, that $0 = s_0 \leq s_1 < s_2 < \ldots < s_p \leq 1$ and let $\alpha_1, \alpha_2, \ldots, \alpha_p$ be real numbers. We will use the Cramér-Wold device and prove that

$$\sum_{u=1}^{p} \alpha_u Z_n^{(b)}(z_u, s_u) \xrightarrow{d} \sum_{u=1}^{p} \alpha_u \mathcal{N}_{z_u, s_u} \quad \text{as} \quad n \rightarrow \infty.$$

We have the following representation:

$$\sum_{u=1}^{p} \alpha_u Z_n^{(b)}(z_u, s_u)$$

$$= \sum_{u=0}^{p-1} \frac{\sqrt{n}}{k} \sum_{j=[ks_u]+1}^{[ks_{u+1}]} \sum_{i=1}^{n} \sum_{v=u+1}^{p} \alpha_v \mathbf{I}(I_{nj} = i)(f_{li}(z_v) - f_n(z_v)). \qquad (31)$$

Denote

$$y_i^u = \sum_{v=u+1}^{p} \alpha_v \mathbf{I}(x_i \leq z_v)$$

and apply Proposition 8.1 with x_i replaced by y_i^u to get, as $n \rightarrow \infty$

$$\frac{\sqrt{n}}{k} \sum_{j=[ks_u]+1}^{[ks_{u+1}]} \sum_{i=1}^{n} \sum_{v=u+1}^{p} \alpha_v \mathbf{I}(I_{nj} = i)(f_{li}(z_v) - f_n(z_v)) \xrightarrow{d} \mathcal{N}(0, (s_{u+1} - s_u)\sigma_u^2),$$

provided the existence of

$$\sigma_u^2 = \lim_{n \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{n} \left(\overline{y_{li}^u} - \overline{y_n^u}\right)^2.$$

Using (28), we get

$$\frac{1}{k}\sum_{i=1}^{n}\left(\overline{y_{li}^u} - \overline{y_n^u}\right)^2 = \frac{1}{k}\sum_{i=1}^{n}\left(\sum_{v=u+1}^{p}\alpha_v\left(f_{li}(z_v) - f_n(z_v)\right)\right)^2$$

$$= \frac{1}{k}\sum_{i=1}^{n}\sum_{v=u+1}^{p}\alpha_v^2\left(f_{li}(z_v) - f_n(z_v)\right)^2$$

$$+ \frac{2}{k}\sum_{i=1}^{n}\sum_{v=u+1}^{p-1}\sum_{v'=v+1}^{p}\alpha_v\alpha_{v'}\left(f_{li}(z_v) - f_n(z_v)\right)\left(f_{li}(z_{v'}) - f_n(z_{v'})\right)$$

$$\rightarrow \sum_{v=u+1}^{p}\alpha_v^2\sigma(z_v, z_v) + 2\sum_{v=u+1}^{p-1}\sum_{v'=v+1}^{p}\alpha_v\alpha_{v'}\sigma(z_v, z_{v'}).$$

Remark now that the u-indexed sums in (31) are independent and

$$
\sum_{u=0}^{p-1}(s_{u+1} - s_u)\sigma_u^2
$$

$$
= \sum_{u=0}^{p-1}(s_{u+1} - s_u)\left[\sum_{v=u+1}^{p} \alpha_v^2 \sigma(z_v, z_v) + 2\sum_{v=u+1}^{p-1}\sum_{v'=v+1}^{p} \alpha_v \alpha_{v'} \sigma(z_v, z_{v'})\right]
$$

$$
= \sum_{v=1}^{p} s_v \alpha_v^2 \sigma(z_v, z_v) + 2\sum_{v=1}^{p-1}\sum_{v'=v+1}^{p} s_v \alpha_v \alpha_{v'} \sigma(z_v, z_{v'})
$$

$$
= Var\left(\sum_{u=1}^{p} \alpha_u \mathcal{N}_{z_u, s_u}\right).
$$

This completes the proof of the proposition. $\square$

Next, we prove tightness of the sequence $(Z_n^{(b)}(\cdot, \cdot))$.

Proposition 8.3. *Let $\{x_i, 1 \le i \le n\}$ and $\{I_{ni}, 1 \le i \le n\}$ be as in Proposition 8.1, and assume that $0 \le x_i \le 1$ for every $0 \le i \le n$. Let k and l be integers such that $n = kl$ and*

$$
l \le C_1 n^{\frac{1}{2} - a}, \quad \text{for some} \quad 0 < a < \frac{1}{2} \quad \text{and} \quad C_1 > 0. \tag{32}
$$

Assume there are constants $C_2 > 0$, $0 < b < 1$ and $c > 0$ such that, for every x and y in $[0, 1]$ and every $n \ge 1$, we have

$$
V_n(x, y) = \frac{1}{k}\sum_{i=1}^{n}(f_{li}(x, y) - f_n(x, y))^2 \le C_2(|x - y|^b + n^{-c}). \tag{33}
$$

Then $(Z_n^{(b)}(\cdot, \cdot))_n$ defined by (22) is tight in $D([0, 1]^2)$; in particular, for every $\varepsilon, \eta > 0$ there exists δ, $0 < \delta < 1$ and N_0 such that for every $n \ge N_0$,

$$
P\left(\sup_{\substack{|x-y|<\delta \\ |s-t|<\delta}} \left|Z_n^{(b)}(x, s) - Z_n^{(b)}(y, t)\right| \ge \varepsilon\right) \le \eta, \tag{34}
$$

and consequently, if Y is taken as a limiting distribution on a subsequence, $P(Y \in C([0, 1]^2)) = 1$.

For computational clarity, we will identify all of the constants involved in the following proof. Note that we now assume $0 \le x_i \le 1$, $i = 1, ..., n$.

Proof. The tightness of the sequence $Z_n^{(b)}(x, s)$ will be proven by closely following the approach used by Naik-Nimbalkar and Rajarshi in [13] and using a restricted chaining argument given in Theorem VII.26 in [15] applied with the

semimetric $d((x,s),(y,t)) = C_3 \max(|x-y|^{\frac{b}{2}}, |s-t|^{\frac{b}{2}})$, where x, y, s, t in $[0,1]$, $0 < b < 1$ and $C_3 > 0$.

By (22), we have the following representation

$$Z_n^{(b)}(x,s) = \sum_{j=1}^{[ks]} E_{nj}(x),$$

where

$$E_{nj}(x) = \frac{\sqrt{n}}{k} \sum_{i=1}^{n} \mathbf{I}(I_{nj} = i)(f_{li}(x) - f_n(x)).$$

Define

$$E_{nj}(x,y) = E_{nj}(x) - E_{nj}(y)$$

and suppose for instance that $s \leq t$, so we can now obtain by virtue of the independence of the E_{nj}'s and inequality (33) that

$$Var\left(Z_n^{(b)}(x,s) - Z_n^{(b)}(y,t)\right)$$

$$= \sum_{j=1}^{[ks]} Var\left(E_{nj}(x,y)\right) + \sum_{j=[ks]+1}^{[kt]} Var\left(E_{nj}(y)\right)$$

$$\leq V_n(x,y) + \left(t - s + \frac{1}{k}\right) V_n(y,0)$$

$$\leq C_2(|x-y|^b + n^{-c}) + C_2\left(|t-s|^b + \frac{1}{k}\right)(1 + n^{-c})$$

$$\leq C_2'\left(|x-y|^b + |t-s|^b + \frac{1}{k} + n^{-c}\right)$$

$$= D_n(x,y,s,t).$$

We also have by (32)

$$|E_{nj}(x,y)| \leq \frac{2\sqrt{n}}{k} \leq 2C_1 n^{-a} \quad \text{and} \quad |E_{nj}(y)| \leq \frac{\sqrt{n}}{k} \leq C_1 n^{-a}.$$

Therefore, by Bennett's inequality (see [15], page 192), we have for every $\eta > 0$

$$P\left(\left|Z_n^{(b)}(x,s) - Z_n^{(b)}(y,t)\right| > \eta\right)$$

$$\leq 2\exp\left(-\frac{1}{2} \frac{\eta^2}{D_n(x,y,s,t)} B\left(\frac{2C_1 n^{-a}\eta}{D_n(x,y,s,t)}\right)\right), \tag{35}$$

where $B(\lambda) = 2\lambda^{-2}[(1+\lambda)\log(1+\lambda) - \lambda]$ for $\lambda > 0$.

Remark now that for any $\delta > 0$

$$d((x,s),(y,t)) \leq \delta \quad \Rightarrow \quad \max(|x-y|,|s-t|) \leq \left(\frac{\delta}{C_3}\right)^{\frac{2}{b}}.$$

Hence, the covering number, which is the smallest m for which there exist points $(x_{\alpha 1}, s_{\alpha 1}), \ldots, (x_{\alpha m}, s_{\alpha m})$ with $\min_i d((x, s), (x_{\alpha i}, s_{\alpha i})) \leq \delta$ for every (x, s) in $[0, 1]^2$, is found to be in this case

$$N(\delta) = N(\delta, d, [0, 1]^2) = \left(\left[\frac{C_3}{\delta} \right]^{\frac{2}{b}} + 1 \right)^2. \tag{36}$$

Let us denote the nearest member of the α-net of $[0, 1]^2$ to (x, s) with respect to the semimetric d by (x_α, s_α). If $d((x, s), (y, t)) \leq \delta$ and $n^{-r} \leq \delta^2$, where $r = \min(\frac{1}{2} + a, c)$, then $D_n(x, y, s, t) \leq C_4 \delta^2$ for some $C_4 > 0$ by (32).

It will be shown later that

i) For every $\lambda \in (0, 1)$, every $\eta > 0$ and $\delta > 0$ such that $\delta^2 \geq n^{-r}$ and $\frac{\delta^2}{\eta} \geq \frac{2C_1 n^{-a}}{C_4 B^{-1}(\lambda)}$, we have

$$P\left(\left| Z_n^{(b)}(x, s) - Z_n^{(b)}(y, t) \right| \geq \eta \right) \leq 2 \exp\left(-\frac{\frac{1}{2}\eta^2 \lambda}{C_4 \delta^2} \right), \tag{37}$$

provided that $d((x, s), (y, t)) \leq \delta$.

ii) For any $\nu > 0$

$$\limsup_{n \to \infty} P\left[\sup_{(x,s) \in [0,1]^2} \left| Z_n^{(b)}(x, s) - Z_n^{(b)}(x_\alpha, s_\alpha) \right| \geq \nu \right] = 0 \text{ a.s,} \tag{38}$$

where $\alpha^2 = \alpha_n^2 = n^{-r} + \frac{2C_1 n^{-a}}{C_4 B^{-1}(\lambda)}$.

To apply Theorem VII.26 of Pollard [15], it remains to show that the associated covering integral with respect to the δ-net of the semimetric d is finite for any $0 < \delta \leq 1$. By definition, the covering integral is

$$\begin{aligned}
J(\delta) = J(\delta, d, [0, 1]^2) &= \int_0^\delta \left(2 \log \left(\frac{N(u)^2}{u} \right) \right)^{\frac{1}{2}} du \\
&\leq \int_0^\delta \left(2 \log \left(\frac{\left(\left(\frac{C}{u}\right)^{\frac{2}{b}} + 1 \right)^4}{u} \right) \right)^{\frac{1}{2}} du \text{ by (36)} \\
&\leq C\delta \left(\log \left(\frac{1}{\delta} \right) \right)^{\frac{1}{2}} \quad \text{for } \delta \text{ small enough.}
\end{aligned}$$

Now, taking $\lambda = 1/4$, $D = 2\sqrt{C_4}$ and $\alpha^2 = n^{-r} + 2C_1 n^{-a}/C_4 B^{-1}(1/4)$ in Theorem VII.26 of Pollard [15], inequality (34) holds.

To complete the proof of the proposition, we shall prove inequalities (37) and (38).

If $d((x, s), (y, t)) \leq \delta$ and $n^{-r} \leq \delta^2$, then $D_n(x, y, s, t) \leq C_4 \delta^2$ as mentioned before. Since $\lambda B(\lambda)$ is an increasing function, we have by (35)

$$P\left(\left|Z_n^{(b)}(x, s) - Z_n^{(b)}(x, s)\right| > \eta\right) \leq 2 \exp\left(-\frac{1}{2}\frac{\eta^2}{C_4 \delta^2} B\left(\frac{2C_1 n^{-a}\eta}{C_4 \delta^2}\right)\right).$$

Further, if $\dfrac{\delta^2}{\eta} \geq \dfrac{2C_1 n^{-a}}{C_4 B^{-1}(\lambda)}$, then inequality (37) holds since $B(0^+) = 1$ and $B(\cdot)$ is a continuous, decreasing function.

We now verify (38). Recall the representation (22)

$$Z_n^{(b)}(x, s) = \sum_{j=1}^{[ks]} E_{nj}(x),$$

where

$$E_{nj}(x) = \frac{\sqrt{n}}{k} \sum_{i=1}^{n} \mathbf{I}(I_{nj} = i)(f_{li}(x) - f_n(x)).$$

Let x and y be any points in $[0, 1]$. Then, arguing as in [13]

$$\begin{aligned}
f_n(x) = f_n(y) &\Rightarrow \mathbf{I}(x_j \leq x) = \mathbf{I}(x_j \leq y) \quad \text{for } j = 1, \ldots, n \\
&\Rightarrow \mathbf{I}(x_{nj} \leq x) = \mathbf{I}(x_{nj} \leq y) \quad \text{for } j = 1, \ldots, n + l - 1 \\
&\Rightarrow Z_n^{(b)}(x, s) = Z_n^{(b)}(y, s) \quad \text{for } s \in [0, 1].
\end{aligned}$$

Since f_n and $[ks]$ assume $(n + 1)$ and $(k + 1)$ different values respectively, $Z_n^{(b)}(x, s) - Z_n^{(b)}(y, t)$ assumes at most $(n + 1)^2 (k + 1)^2$ values as (x, s) and (y, t) vary in $[0, 1]^2$. Therefore,

$$\begin{aligned}
&P\left[\sup_{(x,s) \in [0,1]^2} |Z_n^{(b)}(x, s) - Z_n^{(b)}(x_\alpha, s_\alpha)| > \nu\right] \\
&\leq (n + 1)^2 (k + 1)^2 \sup_{(x,s) \in [0,1]^2} P\left[|Z_n^{(b)}(x, s) - Z_n^{(b)}(x_\alpha, s_\alpha)| > \nu\right].
\end{aligned}$$

Suppose for instance that $s \leq s_\alpha$, then by the kind of computations seen before, we get

$$Var\left(Z_n^{(b)}(x, s) - Z_n^{(b)}(x_\alpha, s_\alpha)\right) \leq D_n(x, x_\alpha, s, s_\alpha).$$

Bernstein's inequality (see [15], page 193) leads to the following:

$$P\left[|Z_n^{(b)}(x, s) - Z_n^{(b)}(x_\alpha, s_\alpha)| > \nu\right] \leq 2 \exp\left(-\frac{\frac{1}{2}\nu^2}{D_n(x, x_\alpha, s, s_\alpha) + \frac{2C_1 n^{-a}}{3}\nu}\right).$$

The definition of α and the condition (32) imply that

$$D_n(x, x_\alpha, s, s_\alpha) + \frac{2C_1 n^{-a}}{3}\nu \leq C_5 n^{-p},$$

for some constants $C_5 > 0$ and $p > 0$. Hence,

$$P\left[\sup_{(x,s)\in[0,1]^2} |Z_n^{(b)}(x,s) - Z_n^{(b)}(x_\alpha, s_\alpha)| > \nu\right] \leq 2(n+1)^4 \exp(-C_6 n^p).$$

This completes the proof of inequality (38) and that of Proposition 8.3. $\quad\square$

We now return to the proof of Theorem 2.2. We will be applying Propositions 8.2 and 8.3 to a fixed trajectory $x_i = X_i(\omega), i = 1, 2, ...$ of the sequence of random variables $\{X_i\}$.

Define the variables $U_i = F(X_i)$ and consider the bootstrapped empirical process $\overline{W}_n^{(b)}(\cdot)$ based on the U_i's, as in (4). Let

$$V_{n1}(x,y) = Var\left(\overline{W}_n^{(b)}(x) - \overline{W}_n^{(b)}(y)\right).$$

Now replace P with P^* and $x_i = U_i(\omega)$ in Propositions 8.2 and 8.3, where P^* denotes the conditional probability given the sample $(X_1, X_2, \cdots, X_n)$. The almost sure convergence of the sequential bootstrapped empirical process $\overline{W}_n^{(b)}(\cdot, \cdot)$ will follow from Propositions 8.2 and 8.3 provided that conditions (27) and (33) hold almost surely: i.e.

$$\lim_{n\to\infty} V_{n1}(x,y) = \sum_{i\in\mathbb{Z}} Cov\left(\mathbf{I}(U_0 \leq x), \mathbf{I}(U_i \leq y)\right) \text{ a.s.}, \tag{39}$$

and for each x and y in $[0,1]$

$$V_{n1}(x,y) \leq C(|x-y|^b + n^{-c}) \quad \text{a.s.}, \tag{40}$$

for some $C > 0$, depending only on trajectory, $0 < b < 1$ and $C > 0$. This was proven by Peligrad under the conditions of our Theorem 2.2 (Proposition 4.1 and proof of Theorem 2.2 in [14]). The conclusion of our Theorem 2.2 can now be derived in a routine manner, as in [2], for the sequential bootstrapped empirical process $W_n^{(b)}(x,s) = \overline{W}_n^{(b)}(F(x), s)$. $\quad\square$

8.2. Proofs of Theorems 3.2 and 3.3

We begin with a slight variation of Theorem 2.2. Let $(x_n)_{n\in\mathbb{Z}}$ be a realization of a stationary sequence $(X_n)_{n\in\mathbb{Z}}$ and let $(\theta_n)_{n\in\mathbb{Z}}$ be a sequence in $[0,1]$ such that $\theta_n \to \theta \in [0,1]$. We introduce some notation: for $1 \leq i \leq [n\theta_n]$ we define $F_{l,i}^{X,\theta_n}$ by

$$F_{l,i}^{X,\theta_n}(x) = \frac{1}{l} \sum_{j=i}^{i+l-1} \mathbf{I}(X_{nj}^* \leq x),$$

where

$$X_{ni}^* = \begin{cases} X_i & \text{if} \quad 1 \leq i \leq [n\theta_n] \\ X_{i-[n\theta_n]} & \text{if} \quad [n\theta_n] + 1 \leq i \leq [n\theta_n] + l - 1. \end{cases} \tag{41}$$

Let

$$W_n^{X,(b,\theta_n)}(x,s) := \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \left[\sum_{i=1}^{[n\theta_n]} \mathbf{I}(I_{nj}=i)(F_{l,i}^{X,\theta_n}(x) - F_{[n\theta_n]}^X(x)) \right], \qquad (42)$$

where $F_{[n\theta_n]}^X(x))$ is the empirical distribution based on $X_1,...,X_{[n\theta_n]}$.

Next, for $[n\theta_n]+1 \le i \le n$, let

$$\tilde{F}_{l,i}^{X,\theta_n}(x) = \frac{1}{l} \sum_{j=i}^{i+l-1} \mathbf{I}(\tilde{X}_{nj}^* \le x),$$

where

$$\tilde{X}_{ni}^* = \begin{cases} X_i & \text{if } [n\theta_n]+1 \le i \le n \\ X_{i-n+[n\theta_n]} & \text{if } n+1 \le i \le n+l-1. \end{cases} \qquad (43)$$

Define

$$\tilde{W}_n^{X,(b,\theta_n)}(x,s) = \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \left[\sum_{i=[n\theta_n]+1}^{n} \mathbf{I}(I_{nj}=i)(\tilde{F}_{l,i}^{(X,\theta_n)}(x) - F_{n-[n\theta_n],[n\theta_n]+1}^X(x)) \right]$$

$$(44)$$

where

$$F_{n-[n\theta_n],[n\theta_n]+1}^X(x) = \frac{1}{n-[n\theta_n]} \sum_{j=[n\theta_n]+1}^{n} \mathbf{I}(X_{nj} \le x).$$

Given a realization $x_1, x_2, \dots$ of $X_1, X_2, \dots$, we use x_{ni}^*, $\tilde{x}_{ni}^*$, $f_{l,i}^{X,\theta_n}(x)$, $f_{n-[n\theta_n],[n\theta_n]+1}^X(x)$, etc. to denote the sample based versions of X_{ni}^*, $\tilde{X}_{ni}^*$, $F_{l,i}^{X,\theta_n}(x)$, $F_{n-[n\theta_n],[n\theta_n]+1}^X(x)$, and so on.

Theorem 8.4. *Assume that the conditions of Theorem 2.2 are satisfied.*
(a) Define $W_n^{X,(b,\theta_n)}(\cdot,\cdot)$ as in (42). If $\theta_n \to \theta > 0$ or if $\theta_n \to 0$ and $l_n\theta_n \to 0$, then as $n \to \infty$,

$$W_n^{X,(b,\theta_n)}(\cdot,\cdot) \xrightarrow{\mathbf{D}} W_\theta^X(\cdot,\cdot)$$

almost surely where $W_\theta^X(\cdot,\cdot)$ is a Gaussian process with zero mean and covariance

$$Cov\left(W_\theta^X(x,s), W_\theta^X(y,t)\right) = \theta(s \wedge t) \sum_{i \in \mathbb{Z}} Cov\left(\mathbf{I}(X_0 \le x), \mathbf{I}(X_i \le y)\right). \qquad (45)$$

(b) Define $\tilde{W}_n^{X,(b,\theta_n)}(\cdot,\cdot)$ as in (44). If $\theta_n \to \theta < 1$ or if $\theta_n \to 1$ and $l_n(1-\theta_n) \to 0$, then as $n \to \infty$,

$$\tilde{W}_n^{X,(b,\theta_n)}(\cdot,\cdot) \xrightarrow{\mathbf{D}} W_{(1-\theta)}^X(\cdot,\cdot)$$

almost surely where $W_{(1-\theta)}^X(\cdot,\cdot)$ is a Gaussian process with zero mean and covariance

$$Cov\left(W_{(1-\theta)}^X(x,s), W_{(1-\theta)}^X(y,t)\right) = (1-\theta)(s \wedge t) \sum_{i \in \mathbb{Z}} Cov\left(\mathbf{I}(X_0 \le x), \mathbf{I}(X_i \le y)\right).$$

$$(46)$$

Furthermore, the limits in (a) and (b) are independent.

We defer the proof of Theorem 8.4 to the end of the subsection.

The key to proving Theorems 3.2 and 3.3 is the following representation of $W_n^{(b)}(x, s)$:

Lemma 8.5. *If* $(X_1, ..., X_n) = (Y_1, ..., Y_{[n\theta_n]}, Z_{[n\theta_n]+1}, ..., Z_n)$, *then*

$$
\begin{aligned}
& W_n^{(b)}(x, s) \\
=\ & W_n^{Y,(b,\theta_n)}(x, s) + \tilde{W}_n^{Z,(b,\theta_n)}(x, s) \quad\quad\quad\quad\quad (47) \\
+\ & \left(F_{[n\theta_n]}^Y(x) - F_{n-[n\theta_n],[n\theta_n]+1}^Z(x) \right) \sqrt{l} B_n^{(\theta_n)}(s) \quad\quad (48) \\
+\ & o_{P^*}(1),
\end{aligned}
$$

where

$$
B_n^{(\theta_n)}(s) = \frac{1}{\sqrt{k}} \sum_{j=1}^{[ks]} \left(\mathbf{I}(I_{nj} \le [n\theta_n]) - \frac{[n\theta_n]}{n} \right)
$$

and the $o_{P^*}(1)$ *term is uniform in* (x, s). *(P^* denotes the conditional probability given the sample* $(X_1, X_2, \cdots, X_n)$*).*

Proof. This can be seen as follows:

$$
\begin{aligned}
W_n^{(b)}(x, s) =\ & \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \left[\sum_{i=1}^n \mathbf{I}(I_{nj} = i)(F_{l,i}^X(x) - F_n^X(x)) \right] \\
=\ & \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \left[\sum_{i=1}^{[n\theta_n]} \mathbf{I}(I_{nj} = i)(F_{l,i}^X(x) - F_{[n\theta_n]}^Y(x)) \right] \\
& + \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \left[\sum_{i=[n\theta_n]+1}^n \mathbf{I}(I_{nj} = i)(F_{l,i}^X(x) - F_{n-[n\theta_n],[n\theta_n]+1}^Z(x)) \right] \\
& + \frac{\sqrt{n}}{k} \left[(F_{[n\theta_n]}^Y(x) - F_n^X(x)) \sum_{j=1}^{[ks]} \mathbf{I}(I_{nj} \le [n\theta_n]) \right. \quad\quad (49) \\
& \left. + (F_{n-[n\theta_n],[n\theta_n]+1}^Z(x) - F_n^X(x)) \sum_{j=1}^{[ks]} (1 - \mathbf{I}(I_{nj} \le [n\theta_n])) \right] \quad (50)
\end{aligned}
$$

We make the following observations:

$$
\begin{aligned}
& \sup_{x,s} \left| \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \sum_{i=1}^{[n\theta_n]} \mathbf{I}(I_{nj} = i)(F_{l,i}^X(x) - F_{l,i}^{Y,\theta_n}(x)) \right| \\
=\ & \sup_{x,s} \left| \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \sum_{i=[n\theta_n]-l+2}^{[n\theta_n]} \mathbf{I}(I_{nj} = i)(F_{l,i}^X(x) - F_{l,i}^{Y,\theta_n}(x)) \right|
\end{aligned}
$$

$$\leq \quad \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \sum_{i=[n\theta_n]-l+2}^{[n\theta_n]} \mathbf{I}(I_{nj}=i) \tag{51}$$

Taking expectations, we see that almost surely,

$$E_{P^*}\left[\sup_{x,s}\left|\frac{\sqrt{n}}{k}\sum_{j=1}^{[ks]}\sum_{i=1}^{[n\theta_n]}\mathbf{I}(I_{nj}=i)(F_{l,i}^X(x)-F_{l,i}^{Y,\theta_n}(x))\right|\right] \leq \frac{\sqrt{n}[ks]l}{kn} \leq \frac{l}{\sqrt{n}},$$

where E_{P^*} denotes conditional expectation with respect to P^*. Therefore,

$$\sup_{x,s}\left|\frac{\sqrt{n}}{k}\sum_{j=1}^{[ks]}\left[\sum_{i=1}^{[n\theta_n]}\mathbf{I}(I_{nj}=i)(F_{l,i}^X(x)-F_{[n\theta_n]}^Y(x))\right]-W_n^{Y,(b,\theta_n)}(x,s)\right|=o_{P^*}(1).$$
$$\tag{52}$$

Likewise,

$$\sup_{x,s}\left|\frac{\sqrt{n}}{k}\sum_{j=1}^{[ks]}\left[\sum_{i=[n\theta_n]+1}^{n}\mathbf{I}(I_{nj}=i)(F_{l,i}^X(x)-F_{n-[n\theta_n],[n\theta_n]+1}^Z(x))\right]-\tilde{W}_n^{Z,(b,\theta_n)}(x,s)\right|$$
$$= o_{P^*}(1). \tag{53}$$

Next, noting that $F_n^X(x)=\frac{[n\theta_n]}{n}F_{[n\theta_n]}^Y(x)+\frac{n-[n\theta_n]}{n}F_{n-[n\theta_n],[n\theta_n]+1}^Z(x)$, a bit of algebra yields

$$(49)+(50)$$
$$= \sqrt{\frac{l}{k}}\left(F_{[n\theta_n]}^Y(x)-F_{n-[n\theta_n],[n\theta_n]+1}^Z(x)\right)\sum_{j=1}^{[ks]}\left(\mathbf{I}(I_{nj}\leq[n\theta_n])-\frac{[n\theta_n]}{n}\right)$$
$$= \left(F_{[n\theta_n]}^Y(x)-F_{n-[n\theta_n],[n\theta_n]+1}^Z(x)\right)\sqrt{l}B_n^{(\theta_n)}(s), \tag{54}$$

where $B_n^{(\theta_n)}(s)=\frac{1}{\sqrt{k}}\sum_{j=1}^{[ks]}\left(\mathbf{I}(I_{nj}\leq[n\theta_n])-\frac{[n\theta_n]}{n}\right)$. Putting (52), (53) and (54) together, we have the desired representation. $\qquad\square$

Proof of Theorem 3.2. Consider the asymptotic behaviour of (48). By ergodicity, since both $[n\theta_n]\to\infty$ and $n-[n\theta_n]\to\infty$, it is immediate that $\left(F_{[n\theta_n]}^Y(\cdot)-F_{n-[n\theta_n],[n\theta_n]+1}^Z(\cdot)\right)\to F(\cdot)-G(\cdot)$ almost surely in $D(\mathbb{R})$ as $n\to\infty$. Next, suppressing dependence of l and k on n for notational convenience and recalling that the I_{nj}'s are i.i.d.,

$$Var\sqrt{l}B_n^{(\theta_n)}(s)=l\frac{[ks]}{k}\frac{[n\theta_n]}{n}\left(1-\frac{[n\theta_n]}{n}\right)\to 0,$$

since $l_n\theta_n(1-\theta_n)\to 0$. Therefore it follows that (48) converges to 0 almost surely in $D(\mathbb{R}\times[0,1])$, provided that $\sqrt{l}B_n^{(\theta_n)}(\cdot)$ is tight in $D[0,1]$. This is an

easy consequence of Theorem 15.6 of [2], since for $s < u < t$ and n sufficiently large that $l\frac{[n\theta_n]}{n}\left(1 - \frac{[n\theta_n]}{n}\right) \leq 1$,

$$E\left[|\sqrt{l}B_n^{(\theta_n)}(u) - \sqrt{l}B_n^{(\theta_n)}(s)|^2 \cdot |\sqrt{l}B_n^{(\theta_n)}(t) - \sqrt{l}B_n^{(\theta_n)}(s)|^2\right]$$

$$= \left(l\frac{[n\theta_n]}{n}\left(1 - \frac{[n\theta_n]}{n}\right)\right)^2 \frac{1}{k^2}([ku] - [ks])([kt] - [ks])$$

$$\leq 4(t - s)^2,$$

where the final inequality follows exactly as in the proof of equation (16.4) in [2]. Therefore, the asymptotic behaviour of $W_n^{(b)}(\cdot, \cdot)$ is the same as that of (47), and from Theorem 8.4 we have

$$W_n^{(b)}(\cdot, \cdot) \overset{\mathbf{D}}{\to} \begin{cases} W_1^Z(\cdot, \cdot) & \text{if } \theta_n \to 0 \\ W_1^Y(\cdot, \cdot) & \text{if } \theta_n \to 1 \end{cases},$$

almost surely as $n \to \infty$. This completes the proof of Theorem 3.2. $\qquad\square$

Proof of Theorem 3.3. We have $\theta_n \to \theta \in (0, 1)$. By Theorem 8.4, we have that the asymptotic behaviour of $\frac{1}{\sqrt{l}}W_n^{(b)}(\cdot, \cdot)$ is almost surely the same as that of $\left(F_{[n\theta_n]}^Y(x) - F_{n-[n\theta_n],[n\theta_n]+1}^Z(x)\right)B_n^{(\theta_n)}(s)$. By the same sort of argument as in the preceding proof, $B_n^{(\theta_n)}(\cdot)$ is tight in $[0, 1]$, and by considering the finite dimensional distributions, it is easily seen that $B_n^{(\theta_n)}(\cdot) \overset{\mathbf{D}}{\to} \sqrt{\theta(1 - \theta)}B(\cdot)$, where $B(\cdot)$ is a standard Brownian motion process on $[0, 1]$. Since $\left(F_{[n\theta_n]}^Y(\cdot) - F_{n-[n\theta_n],[n\theta_n]+1}^Z(\cdot)\right) \overset{\mathbf{D}}{\to} F(\cdot) - G(\cdot)$ almost surely in $D(\mathbb{R})$, it follows that

$$\frac{1}{\sqrt{l}}W_n^{(b)}(\cdot, \cdot) \overset{\mathbf{D}}{\to} L^{(\theta)}(\cdot, \cdot)$$

almost surely in $D(\mathbb{R} \times [0, 1])$, where $L^{(\theta)}(\cdot, \cdot)$ is defined in (15). This completes the proof of Theorem 3.3. $\qquad\square$

It remains only to prove Theorem 8.4.

Proof of Theorem 8.4. We prove (a) only, as the proof of (b) is similar.

Case 1: Suppose that $\theta_n \to 0$ and $l_n\theta_n \to 0$. Recalling that

$$W_n^{X,(b,\theta_n)}(x, s) := \frac{\sqrt{n}}{k}\sum_{j=1}^{[ks]}\left[\sum_{i=1}^{[n\theta_n]}\mathbf{I}(I_{nj} = i)(F_{l,i}^{X,\theta_n}(x) - F_{[n\theta_n]}^X(x))\right], \quad (55)$$

it is straightforward that $E_{P^*}[W_n^{X,(b,\theta_n)}(x, s)] = 0$ and

$$Var_{P^*}([W_n^{X,(b,\theta_n)}(x, s))$$

$$= \frac{n}{k^2} \sum_{j=1}^{[ks]} Var_{P^*} \left(\left[\sum_{i=1}^{[n\theta_n]} \mathbf{I}(I_{nj} = i)(F_{l,i}^{X,\theta_n}(x) - F_{[n\theta_n]}^X(x)) \right] \right)$$

$$= \frac{n}{k^2} \sum_{j=1}^{[ks]} \left[\sum_{i=1}^{[n\theta_n]} P(I_{nj} = i)(F_{l,i}^{X,\theta_n}(x) - F_{[n\theta_n]}^X(x))^2 \right]$$

$$\leq \frac{n}{k^2} [ks] \frac{[n\theta_n]}{n} = O(l_n \theta_n),$$

where $E_{P^*}(\cdot)$ and $Var_{P^*}(\cdot)$ denote conditional mean and variance with respect to P^*. Thus, all the finite dimensional distributions of $W_n^{X,(b,\theta_n)}(\cdot, \cdot)$ converge to 0. The result follows if $W_n^{X,(b,\theta_n)}(\cdot, \cdot)$ is tight, which may be shown exactly as in Case 2, below.

Case 2: Suppose that $\theta_n \to \theta > 0$. The idea is as follows: we extend Propositions 8.1, 8.2 and 8.3 to partial sums where i ranges from 1 to $[n\theta_n]$, and the centering is the average of the first $[n\theta_n]$ terms. To be precise, first we note that Proposition 8.1 can be extended as follows: Under the same conditions on $l = l_n$ and $k = k_n$, defining x_{ni}^* as in (41) and denoting $\overline{x}_{li}^* = \frac{1}{l} \sum_{j=i}^{i+l-1} x_{nj}^*$, if

$$V_{[n\theta_n]} = \frac{1}{k} \sum_{i=1}^{[n\theta_n]} \left(\overline{x}_{li}^* - \overline{x}_{[n\theta_n]} \right)^2 \to \theta \sigma^2 > 0 \quad \text{as} \quad n \to \infty, \tag{56}$$

then, for $0 \leq s < t \leq 1$

$$\frac{\sqrt{n}}{k} \sum_{j=[ks]+1}^{[kt]} \sum_{i=1}^{[n\theta_n]} \mathbf{I}(I_{nj} = i) \left(\overline{x}_{li}^* - \overline{x}_{[n\theta_n]} \right) \xrightarrow{d} \mathcal{N}(0, \theta(t-s)\sigma^2) \quad \text{as} \quad n \to \infty. \tag{57}$$

The proof of (57) follows exactly the proof of Proposition 8.1 by observing that $\frac{1}{k} V_{[n\theta_n]}$ is the variance of

$$\frac{\sqrt{n}}{k} \sum_{i=1}^{[n\theta_n]} \mathbf{I}(I_{nj} = i)(\overline{x}_{li}^* - \overline{x}_{[n\theta_n]}).$$

Similarly, under the conditions of Proposition 8.2 the finite dimensional distributions of

$$Z_n^{(b,\theta_n)}(x,s) = \frac{\sqrt{n}}{k} \sum_{j=1}^{[ks]} \left[\sum_{i=1}^{[n\theta_n]} \mathbf{I}(I_{nj} = i)(f_{l,i}^{X,\theta_n}(x) - f_{[n\theta_n]}^X(x)) \right]. \tag{58}$$

Converge, provided that $\lim_{n\to\infty} V_{[n\theta_n]}(x,y)$ exists, where

$$V_{[n\theta_n]}(x,y) = \frac{1}{k} \sum_{i=1}^{[n\theta_n]} (f_{l,i}^{X,\theta_n}(x,y) - f_{[n\theta_n]}^X(x,y))^2. \tag{59}$$

Proposition 8.3 yields tightness of $Z_n^{(b,\theta_n)}(\cdot, \cdot)$ if (33) is replaced by

$$V_{[n\theta_n]}(x, y) \leq C(|x - y|^b + n^{-c}), \tag{60}$$

$0 < b < 1$ and $c > 0$.

Continuing as in the proof of Theorem 2.2, we may assume without loss of generality that the $X_i's$ are $U(0, 1)$. Replace P with P^* and x_i with $X_i(\omega)$ in the statements above and the result follows provided that

$$\lim_{n \to \infty} V_{[n\theta_n]}(x, y) = \theta \sum_{i \in \mathbb{Z}} Cov\left(\mathbf{I}(X_0 \leq x), \mathbf{I}(X_i \leq y)\right) \text{a.s.}, \tag{61}$$

and for each x and y in $[0, 1]$

$$V_{[n\theta_n]}(x, y) \leq C(|x - y|^b + n^{-c}) \quad \text{a.s.}, \tag{62}$$

for some $C > 0$ depending only on trajectory, $0 < b < 1$ and $c > 0$.

First consider (62). As shown in the proof of Theorem 2.2, for some $C > 0$ depending only on trajectory, $0 < b < 1$ and $c > 0$,

$$
\begin{aligned}
&C(|x - y|^b + n^{-c}) \\
\geq\ &\frac{1}{k} \sum_{i=1}^{n} (F_{l,i}^X(x, y) - F_n^X(x, y))^2 \quad \text{a.s.} \\
\geq\ &\frac{1}{k} \sum_{i=1}^{[n\theta_n]} (F_{l,i}^X(x, y) - F_n^X(x, y))^2 \\
=\ &\frac{1}{k} \sum_{i=1}^{[n\theta_n]} (F_{l,i}^{X,\theta_n}(x, y) - F_n^X(x, y))^2 \\
&+ \frac{1}{k} \sum_{i=[n\theta_n]-l+2}^{[n\theta_n]} (F_{l,i}^X(x, y) - F_{l,i}^{X,\theta_n}(x, y))^2 \\
&- \frac{2}{k} \sum_{i=[n\theta_n]-l+2}^{[n\theta_n]} (F_{l,i}^{X,\theta_n}(x, y) - F_{l,i}^X(x, y))(F_{l,i}^{X,\theta_n}(x, y) - F_n^X(x, y)) \\
\geq\ &\frac{1}{k} \sum_{i=1}^{[n\theta_n]} (F_{l,i}^{X,\theta_n}(x, y) - F_{[n\theta_n]}^X(x, y))^2 - \frac{2l}{k}.
\end{aligned}
$$

Since $l/k \ll n^{-1/3}$, c can be redefined, if necessary, to yield (62).

Next, we consider (61). This follows exactly as in the proof of Proposition 4.1 of [14], replacing summations from 1 to n with summations from 1 to $[n\theta_n]$. The only changes that need to be made are to Peligrad's assumptions (4.7) and (4.8) of [14], which in our context should be, respectively,

$$\frac{Var(\sum_{i=1}^{[n\theta_n]} \mathbf{I}(s < X_i \leq t))}{n} \leq K(|s - t|^b + n^{-d}), \tag{63}$$

for some $K > 0$, $d > 0$, $0 < b < 1$, and

$$\sup_{u,v\in[0,1]} \frac{1}{n} \left| \sum_{i=1}^{[n\theta_n]} \frac{Y_{li}^2(u,v) - EY_{li}^2(u,v)}{l} \right| = O(n^{-g}) \text{ a.s. as } n \to \infty, \qquad (64)$$

for some $g > 0$, where

$$Y_{li}(u,v) = \sum_{j=i}^{i+l-1} \left[I(u < X_j \le v) - (F(v) - F(u)) \right].$$

It is easily seen that (63) follows if (4.7) of [14] holds, where the summation is from 1 to n. This, in turn, is a result of Lemma 4.3 of [14]. To verify (64), we return to the proof of Theorem 2.2 of [14], where Peligrad's assumption (4.8) is verified. A close perusal of the proof indicates that summations from 1 to n can be replaced everywhere by summations from 1 to $[n\theta_n]$, and all the required inequalities remain valid. Since both (63) and (64) are satisfied, (61) follows, and Theorem 8.4(a) is proven.

The independence of the limits in (a) and (b) of Theorem 8.4 follows from the fact that for all $x, y \in \mathbf{R}$ and $s, t \in [0,1]$,

$$Cov_{P^*} \left(W_n^{X,(b,\theta_n)}(x,s), \tilde{W}_n^{X,(b,\theta_n)}(y,t) \right) = 0,$$

where $Cov_{P^*}(\cdot,\cdot)$ denotes conditional covariance with respect to P^*. $\square$

8.3. Proof of Proposition 4.2

As in the proof of Proposition 2.8 of [5], a bit of algebra gives us

$$V_n(x,s) = K_n(x,s) + B_n(x,s),$$

where

$$K_n(x,s) := W_n'(x,s) - \frac{[ns]}{n} W_n'(x,1)$$

$$B_n(x,s) := \sqrt{n} H^{(n)}(x,s) - \frac{[ns]}{\sqrt{n}} H^{(n)}(x,1)$$

$$= \sqrt{n} \left[(s(1-\theta_n)) \wedge (\theta_n(1-s)) \right] (F(x) - G(x))$$

$$+ \frac{ns - [ns]}{\sqrt{n}} H^{(n)}(x,1),$$

and $H^{(n)}(\cdot,\cdot)$ is defined in (13).

By assumption, $K_n(\cdot,\cdot) \xrightarrow{\mathbf{D}} W^{(\theta)}(x,s) - s W^{(\theta)}(x,1)$.

If a sequence $\{\theta_n : n \in \mathbb{N}\}$ of converging alternatives satisfies either 1. or 2. of Proposition 4.2(A), then $\sup_{(x,s)\in\mathbb{R}\times[0,1]} |B_n(x,s)| \xrightarrow{p} \infty$ as $n \to \infty$, and by the continuous mapping theorem it is immediate that $T_i \xrightarrow{p} \infty$, $i = 1, 2$ as $n \to \infty$.

On the other hand, if the alternative is non-converging $(\theta_n \to \theta \in (0,1))$ and Proposition 4.2(B) is satisfied, then we observe that

$$n^{-1/2}V_n(x,s) = n^{-1/2}B_n(x,s) + o_P(1) \xrightarrow{p} (s(1-\theta) \wedge \theta(1-s))(F(x) - G(x)),$$

where convergence in probability is on the space $D(\mathbb{R} \times [0,1])$. The continuous mapping theorem and straightforward calculations complete the proof of (B). □

References

[1] BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics* **5** 1656–1670. MR0383482

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York. MR0233396

[3] DOUKHAN, P. and SURGAILIS, D. (1998). Functional central limit theorem for the empirical process of short memory linear processes. *C.R. Acad. Sci. Paris* **326** 87–92. MR1649521

[4] EL KTAIBI, F. (2015). Asymptotics for the sequential empirical process and testing for distributional change for stationary linear models. Ph.D. thesis, University of Ottawa. http://hdl.handle.net/10393/31916

[5] EL KTAIBI, F. and IVANOFF, B. G. (2016) Change-point detection in the marginal distribution of a linear process. *Electronic Journal of Statistics* **10** 3945–3985. MR3581958

[6] EL KTAIBI, F., IVANOFF, B. G. and WEBER, N.C. (2014). Bootstrapping the empirical distribution of a linear process. *Statistics and Probability Letters* **93** 134–142. MR3244565

[7] GIRAITIS, L., LEIPUS, R. and SURGAILIS, D. (1996). The change-point problem for dependent observations. *Journal of Statistical Planning and Inference* **53** 297–310. MR1407644

[8] INOUE, A. (2001). Testing for the distributional change in time series. *Econometric Theory* **17** 156–187. MR1863569

[9] IVANOFF, B. G. (1980). The function space $D([0,\infty)^q, E)$. *The Canadian Journal of Statistics* **8** 179–191. MR0625374

[10] IVANOFF, B. G. and WEBER, N. C. (2010). Asymptotic results for spatial causal ARMA models. *Electronic Journal of Statistics* **4** 15–35. MR2579552

[11] KÜNSCH, H. R.(1989) The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **17** 1217–1241. MR1015147

[12] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data.* Springer, Berlin. MR2001447

[13] NAIK-NIMBALKAR, U. V. and RAJARSHI, M. B. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations. *Ann. Statis.* **22** (2) 980–994. MR1292552

[14] PELIGRAD, M. (1998). On the blockwise bootstrap for empirical processes for stationary sequences. *The Annals of Probability* **26**(2) 877–901. MR1626543

[15] POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York. MR0762984

[16] RADULOVIĆ, D. (2009). Another look at the disjoint blocks bootstrap. *Test* **18**, 195-212. MR2495971

[17] SHARIPOV, O., TEWES, J. AND WENDLER, M. (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canadian Journal of Statistics* **44**, 300-322. MR3536199

[18] VAN DER VAART, A. AND WELLNER, J. (1996). *Weak Convergence and Empirical Processes.* Springer, New York. MR1385671