JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. (2018)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12431

ORIGINAL ARTICLE

COMBINING CUMULATIVE SUM CHANGE-POINT DETECTION TESTS FOR ASSESSING THE STATIONARITY OF UNIVARIATE TIME SERIES

AXEL BÜCHER,[a] JEAN-DAVID FERMANIAN[b] AND IVAN KOJADINOVIC[c*]

[a] *Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany*
[b] *CREST-ENSAE, Finance and Insurance department, Palaiseau, France*
[c] *Laboratoire de mathématiques et applications – IPRA, CNRS/Université de Pau et des Pays de l'Adour/E2S UPPA, UMR 5142, Pau, France*

We derive tests of stationarity for univariate time series by combining change-point tests sensitive to changes in the contemporary distribution with tests sensitive to changes in the serial dependence. The proposed approach relies on a general procedure for combining dependent tests based on resampling. After proving the asymptotic validity of the combining procedure under the conjunction of null hypotheses and investigating its consistency, we study rank-based tests of stationarity by combining cumulative sum change-point tests based on the contemporary empirical distribution function and on the empirical autocopula at a given lag. Extensions based on tests solely focusing on second-order characteristics are proposed next. The finite-sample behaviors of all the derived statistical procedures for assessing stationarity are investigated in large-scale Monte Carlo experiments, and illustrations on two real datasets are provided. Extensions to multi-variate time series are briefly discussed as well.

Received 08 September 2017; Accepted 22 September 2018

Keywords: Copula; dependent *p*-value combination; multiplier bootstrap; rank-based statistics; tests of stationarity.

MOS subject classification: 62E20; 62G10; 62G09.

1. INTRODUCTION

Testing the stationarity of a time series is of great importance prior to any modeling. Existing approaches assessing whether a time series is stationary could roughly be grouped into two main categories: procedures that mostly work in the frequency domain and those that mostly work in the time domain. Among the tests in the former group, one finds, for instance, approaches testing the constancy of a spectral functional (see, e.g., Priestley and Subba Rao, 1969; Paparoditis, 2010), procedures comparing a time-varying spectral density estimate with its stationary approximation (see, e.g., Dette et al., 2011; Preuss et al., 2013; Puchstein and Preuss, 2016) and approaches based on wavelets (see, e.g., von Sachs and Neumann, 2000; Nason, 2013; Cardinali and Nason, 2013, 2016). As far as the second category of tests is concerned, one mostly finds approaches based on the autocovariance/autocorrelation function, such as those used in Lee *et al.* (2003), Dwivedi and Subba Rao (2011), Jin *et al.* (2015) and Dette *et al.* (2015). In particular, the works of Lee *et al.* (2003) and Dette *et al.* (2015) also clearly pertain to the change-point detection literature (see, e.g., Csörgö and Horváth, 1997; Aue and Horváth, 2013 for an overview). The latter should not come as a surprise. Indeed, any test for change-point detection may be seen as a test of stationarity designed to be sensitive to a particular type of departure from stationarity.

To illustrate the latter point, let $X_1, X_2, \dots$ be a stretch from a univariate time series and consider the classical *cumulative sum* (CUSUM) test 'for a change in the mean' (see, e.g., Page, 1954; Phillips, 1987). The latter is

* Correspondence to: Ivan Kojadinovic, Laboratoire de mathématiques et applications - IPRA, CNRS/Université de Pau et des Pays de l'Adour/E2S UPPA, UMR 5142, B.P. 1155, 64013 Pau, France. E-mail: ivan.kojadinovic@univ-pau.fr

© 2018 John Wiley & Sons Ltd

usually regarded as a test of

$$H_0 : X_1, X_2, \dots \text{ have the same expectation,}$$

but it only holds its level asymptotically if $X_1, X_2, \dots$ is a stretch from a time series whose autocovariances at all lags are constant (Zhou, 2013). Without the latter assumption, a small p-value can only be used to conclude that $X_1, X_2, \dots$ is not a stretch from a second-order stationary time series. In other words, without the additional assumption of constant autocovariances, the classical CUSUM test 'for a change in the mean' is merely a test of second-order stationarity that is particularly sensitive to a change in the expectation.

Obtaining a large p-value when carrying out the previously mentioned test should clearly not be interpreted as no evidence against second-order stationarity since a change in mean is only one possible departure from second-order stationarity. Following Dette *et al.* (2015), complementing the previous test by tests for change-point detection particularly sensitive to changes in the variance and in the autocorrelation at some fixed lags may, in case of large p-values, comfort a practitioner in considering that $X_1, X_2, \dots$ might well be a stretch from a second-order stationary time series. The aim of this work is to adopt a similar perspective on assessing stationarity but without only restricting the analysis to second-order characteristics. In fact, all finite dimensional distributions induced by a time series could be potentially tested.

More formally, suppose we observe a stretch $X_1, \dots, X_N$ from a time series of univariate continuous random variables. For some $2 \leq h \leq N$, set $n = N - h + 1$ and let $\boldsymbol{Y}_1^{(h)}, \dots, \boldsymbol{Y}_n^{(h)}$ be h-dimensional random vectors defined by

$$\boldsymbol{Y}_i^{(h)} = (X_i, \dots, X_{i+h-1}), \qquad i \in \{1, \dots, n\}. \tag{1.1}$$

Note that the quantity h is sometimes called the *embedding dimension*, and $h - 1$ can be interpreted as the maximum lag under investigation. As an imperfect alternative, we shall focus on tests particularly sensitive to departures from the hypothesis

$$H_0^{(h)} : \exists F^{(h)} \text{ such that } \boldsymbol{Y}_1^{(h)}, \dots, \boldsymbol{Y}_n^{(h)} \text{ have the distribution function (d.f.) } F^{(h)}. \tag{1.2}$$

To derive such tests, a first natural approach would be to apply to the random vectors in (1.1) non-parametric CUSUM tests, such as those based on differences of empirical d.f.s studied in Gombay and Horváth (1999), Inoue (2001) and Holmes *et al.* (2013) (see also Section 3.2 below) or on differences of empirical characteristic functions (see, e.g., Hušková and Meintanis, 2006a, 2006b). However, preliminary numerical experiments (some of which are reported in Section 5) revealed the low power of such an adaptation in the case of the empirical d.f.-based tests, especially when the non-stationarity of the underlying univariate time series is a consequence of changes in the serial dependence. These empirical conclusions, in line with those drawn in Bücher *et al.* (2014) in a related context, prompted us to consider the alternative approach consisting of assessing changes in the 'contemporary' distribution (i.e., of the X_i) separately from changes in the serial dependence.

Suppose that $H_0^{(h)}$ in (1.2) holds, and recall that $X_1, \dots, X_{n+h-1}$ is assumed to be a stretch from a time series of univariate continuous random variables. Then, the common d.f. of $\boldsymbol{Y}_i^{(h)}$ can be written (Sklar, 1959) as

$$F^{(h)}(\boldsymbol{x}) = C^{(h)}\{G(x_1), \dots, G(x_h)\}, \quad \boldsymbol{x} \in \mathbb{R}^h,$$

where $C^{(h)}$ is the unique *copula* (merely an h-dimensional d.f. with standard uniform margins) associated with $F^{(h)}$, and G is the common marginal univariate d.f. of all the components of the $\boldsymbol{Y}_i^{(h)}$, $i \in \{1, \dots, n\}$. The copula $C^{(h)}$ controls the dependence between the components of the $\boldsymbol{Y}_i^{(h)}$. Equivalently, it controls the *serial dependence* up to lag $h - 1$ in the time series, which is why it is sometimes called the lag $h - 1$ *serial copula* or *autocopula* in the literature.

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

Notice further that, slightly abusing notation, the hypothesis $H_0^{(h)}$ in (1.2) can be written as $H_0^{(1)} \cap H_{0,c}^{(h)}$, where

$$H_0^{(1)} \,:\, \exists\, G \text{ such that } X_1, X_2, \ldots \text{ have the d.f. } G, \tag{1.3}$$

and

$$H_{0,c}^{(h)} \,:\, \exists\, C^{(h)} \text{ such that } \boldsymbol{Y}_1^{(h)}, \ldots, \boldsymbol{Y}_n^{(h)} \text{ have the copula } C^{(h)}. \tag{1.4}$$

In other words, $H_0^{(h)}$ in (1.2) holds if all the X_i have the same (contemporary) distribution and if all the $\boldsymbol{Y}_i^{(h)}$ have the same copula.

A sensible strategy for assessing whether $H_0^{(h)}$ in (1.2) is plausible would thus naturally consist of combining two tests: a test particularly sensitive to departures from $H_0^{(1)}$ in (1.3) and a test particularly sensitive to departures from $H_{0,c}^{(h)}$ in (1.4). For the former, as already mentioned, a natural candidate in the general context under consideration is the CUSUM test based on differences of empirical d.f.s studied in Gombay and Horváth (1999) and Holmes *et al.* (2013). We shall briefly revisit the latter approach in the setting of serially dependent observations. One of the main goals of this work is to derive a test that is particularly sensitive to departures from $H_{0,c}^{(h)}$ in (1.4), that is, to changes in the serial dependence. The idea is not new but seems to have been used only with respect to second-order characteristics of a time series; see, for example, Lee *et al.* (2003) for tests on the autocovariance in a CUSUM setting and Dwivedi and Subba Rao (2011) and Jin *et al.* (2015) for tests in a different setting. Specifically, one of the main contributions of this work is to propose a CUSUM test that is sensitive to departures from $H_{0,c}^{(h)}$. It will be based on a serial version of the so-called *empirical copula* that we should naturally refer to as the *empirical autocopula* hereafter.

Because the aforementioned test based on empirical d.f.s (particularly sensitive to departures from $H_0^{(1)}$ in (1.3) by construction) and the test based on empirical autocopulas (designed to be sensitive to departures from $H_{0,c}^{(h)}$ in (1.4)) rely on the same type of resampling, bootstrap replicates on the underlying statistics $S_{n,G}$ and $S_{n,C^{(h)}}$ can be generated jointly to reproduce, approximately, the distribution of $(S_{n,G}, S_{n,C^{(h)}})$ under stationarity. Under such an assumption, another main contribution of this work, which may be of independent interest, is a general procedure for combining dependent bootstrap-based tests, relying on appropriate extensions of well-known p-value combination methods such as those of Fisher (1932) or Stouffer *et al.* (1949).

An interesting and desirable feature of the resulting global testing procedure is that it is based on ranks. It is therefore expected to be quite robust in the presence of heavy-tailed observations. Still, in the case of Gaussian time series, some tests based on second-order characteristics might be more powerful. A natural competitor to our aforementioned global test could thus be obtained by combining tests particularly sensitive to changes in the expectation, variance and autocovariances up to lag $h-1$. Interestingly enough, CUSUM versions of such tests can be cast in the setting considered in Bücher and Kojadinovic (2016a): they can all be carried out using the same type of resampling, and thus, as described in the previous paragraph, their (dependent) p-values can be combined, leading to a test that could be regarded as a test of second-order stationarity.

This article is organized as follows. The proposed procedure for combining dependent bootstrap-based tests is described in Section 2, conditions under which it is asymptotically valid under the conjunction of the component null hypotheses are stated, and its consistency is theoretically investigated. The detailed description of the combined rank-based test involving empirical d.f.s and empirical autocopulas is given in Section 3, along with theoretical results about its asymptotic validity under the null hypothesis of stationarity. The choice of the embedding dimension h is discussed in Section 3.4. The fourth section is devoted to related combined tests based on second-order characteristics: the corresponding testing procedures are provided, and asymptotic validity results under the null are stated. Section 5 reports Monte Carlo experiments that are used to empirically study the previously described tests. Some illustrations on real-world data are presented in Section 6. Finally, concluding remarks are provided in Section 7, one of which particularly discusses multi-variate extensions of the proposed tests.

Auxiliary results and all proofs are deferred to a sequence of appendices. Additional theoretical and simulation results are provided in Appendix S1 (Supporting Information). The studied tests are implemented in the package

DOI: 10.1111/jtsa.12431
© 2018 John Wiley & Sons Ltd

npcp (Kojadinovic, 2017) for the R statistical system (R Core Team, 2017). In the rest of the paper, the arrow '$\rightsquigarrow$' denotes weak convergence in the sense of Definition 1.3.3 in van der Vaart and Wellner (2000), while the arrow '$\overset{\mathbb{P}}{\to}$' denotes convergence in probability. All convergences are for $n \to \infty$ if not mentioned otherwise. Finally, given a set S, $\ell^\infty(S)$ denotes the space of all bounded real-valued functions on S equipped with the uniform metric.

2. A GENERAL PROCEDURE TO COMBINE DEPENDENT TESTS BASED ON RESAMPLING

As argued in the introduction, to assess whether stationarity is likely to hold, it might be beneficial to combine several tests, each of which is designed to be sensitive to a particular form of non-stationarity. As the need for similar approaches may arise in other contexts than stationarity testing, in this section, we propose a very general strategy for combining tests based on resampling by relying on well-known p-value combination methods such as those of Fisher (1932) or Stouffer *et al.* (1949). Recall that, given r p-values $p_1, \dots, p_r$ for right-tailed tests of corresponding null hypotheses $H_0^{(1)}, \dots, H_0^{(r)}$ with corresponding strictly positive weights $w_1, \dots, w_r$ that quantify the importance of each test in the combination, the latter method consists of computing, up to a rescaling term, the global statistic

$$\psi_S(p_1, \dots, p_r) = \sum_{j=1}^{r} w_j \Phi^{-1}(1 - p_j), \tag{2.1}$$

where Φ^{-1} is the quantile function of the standard normal. Large values provide evidence against the global null hypothesis $H_0 = H_0^{(1)} \cap \dots \cap H_0^{(r)}$. By analogy, the corresponding weighted version of the global statistic in Fisher's p-value combination method can be defined by

$$\psi_F(p_1, \dots, p_r) = -2 \sum_{j=1}^{r} w_j \log(p_j). \tag{2.2}$$

If the p-values $p_1, \dots, p_r$ are independent and uniformly distributed on $(0, 1)$, then it can be verified that $\psi_S(p_1, \dots, p_r)$ or $\psi_F(p_1, \dots, p_r)$ are pivotal, giving rise to simple exact global tests. If the component tests are dependent, however, the distributions of the previous statistics are not pivotal, and computing the corresponding global p-values is not straightforward anymore.

Let $\boldsymbol{X}_n$ denote the available data (apart from measurability, no assumptions are made on $\boldsymbol{X}_n$, but it is instructive to think of $\boldsymbol{X}_n$ as an n-tuple of possibly multi-variate observations, which may be serially dependent), and let $T_{n,1} = T_{n,1}(\boldsymbol{X}_n), \dots, T_{n,r} = T_{n,r}(\boldsymbol{X}_n)$ be the statistics, each $\mathbb{R}$-valued, of the r tests to be combined.

We assume furthermore that, for any $j \in \{1, \dots, r\}$, large values of $T_{n,j}$ provide evidence against the hypothesis $H_0^{(j)}$. As we continue, we let $\boldsymbol{T}_n = \boldsymbol{T}_n(\boldsymbol{X}_n)$ denote the r-dimensional random vector $(T_{n,1}, \dots, T_{n,r}) = (T_{n,1}(\boldsymbol{X}_n), \dots, T_{n,r}(\boldsymbol{X}_n))$.

We suppose, in addition, that we have available a resampling mechanism that allows us to obtain a sample of M bootstrap replicates $\boldsymbol{T}_n^{[i]} = \boldsymbol{T}_n^{[i]}(\boldsymbol{X}_n, \boldsymbol{V}_n^{[i]})$, $i \in \{1, \dots, M\}$, of $\boldsymbol{T}_n$, where $\boldsymbol{V}_n^{[1]}, \dots, \boldsymbol{V}_n^{[M]}$ are i.i.d. $\mathbb{R}^n$-valued random vectors representing the additional sources of randomness involved in the resampling mechanism such that, for any $i \in \{1, \dots, M\}$, $T_{n,j}^{[i]}$ depends on the data $\boldsymbol{X}_n$ and $\boldsymbol{V}_n^{[i]}$, that is, $T_{n,j}^{[i]} = T_{n,j}^{[i]}(\boldsymbol{X}_n, \boldsymbol{V}_n^{[i]})$ for all $j \in \{1, \dots, r\}$. Note that the previous setup naturally implies that the components $T_{n,1}^{[i]}, \dots, T_{n,r}^{[i]}$ of $\boldsymbol{T}_n^{[i]}$ are bootstrap replicates of the components $T_{n,1}, \dots, T_{n,r}$ of $\boldsymbol{T}_n$. The fact that all the components of $\boldsymbol{T}_n^{[i]}$ depend on the same additional source of randomness $\boldsymbol{V}_n^{[i]}$ makes it possible to expect that the bootstrap replicates $\boldsymbol{T}_n^{[i]}$, $i \in \{1, \dots, M\}$, be, approximately, i.i.d. copies of $\boldsymbol{T}_n$ under the global null hypothesis $H_0 = H_0^{(1)} \cap \dots \cap H_0^{(r)}$. For the individual test based on $T_{n,j}$, $j \in \{1, \dots, r\}$, an approximate p-value could then naturally be computed as

$$\frac{1}{M} \sum_{i=1}^{M} \mathbf{1}(T_{n,j}^{[i]} \geq T_{n,j}).$$

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

Let ψ be a continuous function from $(0,1)^r$ to $\mathbb{R}$ that is decreasing in each of its r arguments (such as ψ_S or ψ_F in (2.1) and (2.2) respectively). To compute an approximate p-value for the global statistic $\psi\{p_{n,M}(T_{n,1}),\dots,p_{n,M}(T_{n,r})\}$, we propose the following procedure:

1. Let $\boldsymbol{T}_n^{[0]} = \boldsymbol{T}_n$.
2. Given a large integer M, compute the sample of M bootstrap replicates $\boldsymbol{T}_n^{[1]},\dots,\boldsymbol{T}_n^{[M]}$ of $\boldsymbol{T}_n^{[0]}$.
3. Then, for all $i \in \{0,1,\dots,M\}$ and $j \in \{1,\dots,r\}$, compute

$$p_{n,M}\left(T_{n,j}^{[i]}\right) = \frac{1}{M+1}\left\{\frac{1}{2} + \sum_{k=1}^{M} \mathbf{1}\left(T_{n,j}^{[k]} \geq T_{n,j}^{[i]}\right)\right\}. \tag{2.3}$$

4. Next, for all $i \in \{0,1,\dots,M\}$, compute

$$W_{n,M}^{[i]} = \psi\left\{p_{n,M}\left(T_{n,1}^{[i]}\right),\dots,p_{n,M}\left(T_{n,r}^{[i]}\right)\right\}. \tag{2.4}$$

5. The global statistic is $W_{n,M}^{[0]}$, and the corresponding approximate p-value is given by

$$p_{n,M}\left(W_{n,M}^{[0]}\right) = \frac{1}{M}\sum_{k=1}^{M} \mathbf{1}\left(W_{n,M}^{[k]} \geq W_{n,M}^{[0]}\right). \tag{2.5}$$

Note that the quantities $p_{n,M}(T_{n,j}^{[i]})$, $j \in \{1,\dots,r\}$, in Step 3, can be regarded as approximate p-values for the 'statistic values' $T_{n,j}^{[i]}$, $j \in \{1,\dots,r\}$. The offset by $1/2$ and the division by $M+1$ instead of M in the formula is carried out to ensure that $p_{n,M}(T_{n,j}^{[i]})$ belongs to the interval $(0,1)$ so that Step 4 is well-defined.

The next result, proven in Appendix B, provides conditions under which the global test based on $W_{n,M}^{[0]}$ given by (2.4) is asymptotically valid under the global null hypothesis $H_0 = H_0^{(1)} \cap \cdots \cap H_0^{(r)}$ and the natural assumption that $M = M_n \to \infty$ as $n \to \infty$. Before proceeding, note that $W_{n,M_n}^{[0]}$ is a Monte Carlo approximation of the unobservable statistic

$$W_n = \psi\left\{\mathbb{P}\left(T_{n,1}^{[1]} \geq T_{n,1} \mid \boldsymbol{X}_n\right),\dots,\mathbb{P}\left(T_{n,r}^{[1]} \geq T_{n,r} \mid \boldsymbol{X}_n\right)\right\}. \tag{2.6}$$

Proposition 2.1. Let $M = M_n \to \infty$ as $n \to \infty$. Assume that $H_0 = H_0^{(1)} \cap \cdots \cap H_0^{(r)}$ holds, that $\boldsymbol{T}_n$ converges weakly to $\boldsymbol{T} = (T_1,\dots,T_r)$, where $\boldsymbol{T}$ has a continuous d.f., and that either

$$\left(\boldsymbol{T}_n, \boldsymbol{T}_n^{[1]}, \boldsymbol{T}_n^{[2]}\right) \rightsquigarrow \left(\boldsymbol{T}, \boldsymbol{T}^{[1]}, \boldsymbol{T}^{[2]}\right), \tag{2.7}$$

where $\boldsymbol{T}^{[1]}$ and $\boldsymbol{T}^{[2]}$ are independent copies of $\boldsymbol{T}$, or

$$\sup_{\boldsymbol{x}\in\mathbb{R}^r}\left|\mathbb{P}\left(\boldsymbol{T}_n^{[1]} \leq \boldsymbol{x} \mid \boldsymbol{X}_n\right) - \mathbb{P}\left(\boldsymbol{T}_n \leq \boldsymbol{x}\right)\right| \overset{\mathbb{P}}{\to} 0. \tag{2.8}$$

Then, for any $N \in \mathbb{N}$,

$$\left(W_{n,M_n}^{[0]}, W_{n,M_n}^{[1]},\dots,W_{n,M_n}^{[N]}\right) \rightsquigarrow \left(W, W^{[1]},\dots,W^{[N]}\right), \tag{2.9}$$

where

$$W = \psi\{\bar{F}_{T_1}(T_1),\dots,\bar{F}_{T_r}(T_r)\} \tag{2.10}$$

is the weak limit of W_n in (2.6) with $\bar{F}_{T_j}(x) = \mathbb{P}(T_j \geq x)$, $x \in \mathbb{R}$, $j \in \{1, \dots, r\}$, and $W^{[1]}, \dots, W^{[N]}$ are independent copies of W. Furthermore, if ψ is chosen in such a way that the random variable W has a continuous d.f., then

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(W^{[1]}_{n,M_n} \leq x \mid \boldsymbol{X}_n \right) - \mathbb{P}(W_n \leq x) \right| \overset{\mathbb{P}}{\to} 0, \tag{2.11}$$

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(W^{[i]}_{n,M_n} \leq x \right) - \mathbb{P}(W_n \leq x) \right| \overset{\mathbb{P}}{\to} 0, \tag{2.12}$$

and, as a consequence, $p_{n,M_n}(W^{[0]}_{n,M_n}) \rightsquigarrow \text{Uniform}(0, 1)$, where $p_{n,M_n}(W^{[0]}_{n,M_n})$ is defined by (2.5).

It is worthwhile mentioning that, by Lemma 2.2 of Bücher and Kojadinovic (2018) and the assumption of continuity for the d.f. of $\boldsymbol{T}$, the statements (2.7) and (2.8) are actually equivalent in the setting under consideration. Notice also that the resulting unconditional bootstrap consistency statement in (2.9) does not require W in (2.10) to have a continuous d.f. Proving the latter might actually be quite complicated as shall be illustrated in a particular case in Section 3.3.

We end this section by providing a result, proven in Appendix B, that states conditions under which the global test based on $W^{[0]}_{n,M}$ given by (2.4) leads to the rejection of the global null hypothesis $H_0 = H_0^{(1)} \cap \cdots \cap H_0^{(r)}$.

Proposition 2.2. Let $M = M_n \to \infty$ as $n \to \infty$. Assume that

(i) the combining function ψ is of the form

$$\psi(p_1, \dots, p_r) = \sum_{j=1}^{r} w_j \varphi(p_j),$$

where φ is decreasing, non-negative and one-to-one from $(0, 1)$ to $(0, \infty)$;

(ii) there exists $j_0 \in \{1, \dots, r\}$ such that the null hypothesis $H_0^{(j_0)}$ of j_0th test T_{n,j_0} does not hold, and $\mathbb{P}\left(T^{[1]}_{n,j_0} \geq T_{n,j_0}\right)$ converges to zero;

(iii) for any $j \in \{1, \dots, r\}$, the sample of bootstrap replicates $T^{[1]}_{n,j}, \dots, T^{[M_n]}_{n,j}$ does not contain ties.

Then, the approximate p-value $p_{n,M_n}\left(W^{[0]}_{n,M_n}\right)$ of the global test converges to zero in probability, where $p_{n,M_n}\left(W^{[0]}_{n,M_n}\right)$ is defined by (2.5).

Let us comment on the assumptions of the previous proposition. Assumption (*i*) is satisfied by the function ψ_F defined by (2.2) but not by the function ψ_S defined by (2.1). A result similar to Proposition 2.2, which can be used to handle the function ψ_S, is stated and proven in Appendix S1. Assumption (*ii*) can, for instance, be shown to hold under the hypothesis of one change in the contemporary d.f. of a time series when T_{n,j_0} is a test statistic such as the one to be defined in Section 3.2, the observations are i.i.d., and the underlying resampling mechanism is a particular multiplier bootstrap. Specifically, in that case, one can rely on Theorem 3 of Holmes *et al.* (2013) to show that, under the hypothesis of one change in the contemporary d.f., T_{n,j_0} diverges to infinity in probability, while $T^{[1]}_{n,j_0}$ is bounded in probability, implying that $T^{[1]}_{n,j_0} - T_{n,j_0}$ diverges to $-\infty$ in probability and, thus, that $\mathbb{P}\left(T^{[1]}_{n,j_0} \geq T_{n,j_0}\right)$ converges to zero. Finally, assumption (*iii*) appears to be empirically satisfied for most bootstrap-based tests for time series of continuous random variables.

3. A RANK-BASED COMBINED TEST SENSITIVE TO DEPARTURES FROM $H_0^{(H)}$

The aim of this section is to use the results of the previous section to derive a global test of stationarity by combining a test that is particularly sensitive to departures from $H_0^{(1)}$ in (1.3) with a test that is particularly sensitive to

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

departures from $H_{0,c}^{(h)}$ in (1.4). We start by describing the latter test and provide conditions under which it is asymptotically valid under stationarity. The available data, denoted generically by $\boldsymbol{X}_n$ in Section 2, take here, as in the introduction, the form of a stretch $X_1, \ldots, X_{n+h-1}$ from a univariate time series, where h is the chosen embedding dimension and where each X_i is assumed to have a continuous d.f.

3.1. A copula-based test sensitive to changes in the serial dependence

The test that we consider has the potential of being sensitive to all types of changes in the serial dependence up to lag $h-1$. Under $H_0^{(h)}$ in (1.2), this serial dependence is completely characterized by the (auto)copula $C^{(h)}$ in (1.4). It is then natural to base the test on *empirical (auto)copulas* (see, e.g., Deheuvels, 1979, 1981) calculated from portions of the data. For any $1 \leq k \leq l \leq n$, let

$$C_{k:l}^{(h)}(\boldsymbol{u}) = \frac{1}{l-k+1} \sum_{i=k}^{l} \prod_{j=1}^{h} \mathbf{1}\{G_{k:l}(X_{i+j-1}) \leq u_j\}, \qquad \boldsymbol{u} \in [0,1]^h, \tag{3.1}$$

where

$$G_{k:l}(x) = \frac{1}{l+h-k} \sum_{j=k}^{l+h-1} \mathbf{1}(X_j \leq x), \qquad x \in \mathbb{R}, \tag{3.2}$$

with the convention that $C_{k:l}^{(h)} = 0$ if $k > l$. The quantity $C_{k:l}^{(h)}$ is a non-parametric estimator of $C^{(h)}$ based on $\boldsymbol{Y}_k^{(h)}, \ldots, \boldsymbol{Y}_l^{(h)}$ that, as already mentioned, we shall call the *lag* $h-1$ *empirical autocopula*. The latter was, for instance, used in Genest and Rémillard (2004) for testing serial independence. It can be verified that it is a straightforward transposition of one of the usual definitions of the empirical copula (when computed from a subsample) to the serial context under consideration.

3.1.1. Test statistic

The CUSUM statistic that we consider is

$$S_{n,C^{(h)}} = \sup_{s\in[0,1]} \int_{[0,1]^h} \left\{\mathbb{D}_{n,C^{(h)}}(s,\boldsymbol{u})\right\}^2 \mathrm{d}C_{1:n}^{(h)}(\boldsymbol{u}) = \max_{1\leq k\leq n-1} \int_{[0,1]^h} \left\{\mathbb{D}_{n,C^{(h)}}(k/n,\boldsymbol{u})\right\}^2 \mathrm{d}C_{1:n}^{(h)}(\boldsymbol{u}), \tag{3.3}$$

where, as mentioned earlier, $\lfloor . \rfloor$ is the floor function,

$$\mathbb{D}_{n,C^{(h)}}(s,\boldsymbol{u}) = \sqrt{n}\lambda_n(0,s)\lambda_n(s,1)\left\{C_{1:\lfloor ns\rfloor}^{(h)}(\boldsymbol{u}) - C_{\lfloor ns\rfloor+1:n}^{(h)}(\boldsymbol{u})\right\}, \qquad (s,\boldsymbol{u}) \in [0,1]^{h+1}, \tag{3.4}$$

and $\lambda_n(s,t) = (\lfloor nt\rfloor - \lfloor ns\rfloor)/n$, $(s,t) \in \Delta = \{(s,t) \in [0,1]^2 : s \leq t\}$.

Under $H_0^{(h)}$ in (1.2), the difference between $C_{1:k}^{(h)}$ and $C_{k+1:n}^{(h)}$ should be small for all $k \in \{1, \ldots, n-1\}$, resulting in small values of $S_{n,C^{(h)}}$. On the contrary, large values of $S_{n,C^{(h)}}$ provide evidence of non-stationarity. The coefficient $\sqrt{n}\lambda_n(0,s)\lambda_n(s,1)$ in (3.4) is the classical normalizing term in the CUSUM approach. It ensures that, under suitable conditions, $S_{n,C^{(h)}}$ converges in distribution under the null hypothesis of stationarity. Analogous to what was explained in the introduction, the test based on $S_{n,C^{(h)}}$ should, in general, not be used to reject $H_{0,c}^{(h)}$ in (1.4). It is merely a test of stationarity that is particularly sensitive to a change in the lag $h-1$ autocopula.

3.1.2. Limiting null distribution

The limiting null distribution of $S_{n,C^{(h)}}$ turns out to be a corollary of a recent result by Bücher and Kojadinovic (2016b) and Bücher *et al.* (2014). Under $H_0^{(h)}$ in (1.2), it can be verified that $\mathbb{D}_{n,C^{(h)}}$ in (3.4) can be written as

$$\mathbb{D}_{n,C^{(h)}}(s,\boldsymbol{u}) = \lambda_n(s,1)\,\mathbb{C}_{n,C^{(h)}}(0,s,\boldsymbol{u}) - \lambda_n(0,s)\,\mathbb{C}_{n,C^{(h)}}(s,1,\boldsymbol{u}), \qquad (s,\boldsymbol{u}) \in [0,1]^{h+1}, \tag{3.5}$$

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

where

$$\mathbb{C}_{n,C^{(h)}}(s,t,\boldsymbol{u}) = \sqrt{n}\,\lambda_n(s,t)\left\{C^{(h)}_{\lfloor ns\rfloor+1:\lfloor nt\rfloor}(\boldsymbol{u}) - C^{(h)}(\boldsymbol{u})\right\}, \qquad (s,t,\boldsymbol{u}) \in \Delta\times[0,1]^h. \tag{3.6}$$

Hence, the null weak limit of the empirical process $\mathbb{D}_{n,C^{(h)}}$ follows from that of $\mathbb{C}_{n,C^{(h)}}$, which we shall call the *sequential empirical autocopula process*.

The following usual condition on the partial derivatives of $C^{(h)}$ (see Segers, 2012) is considered as we continue.

Condition 3.1. For any $j \in \{1,\dots,h\}$, the partial derivative $\dot{C}^{(h)}_j = \partial C^{(h)}/\partial u_j$ exists and is continuous on $V^{(h)}_j = \{\boldsymbol{u} \in [0,1]^h : u_j \in (0,1)\}$.

Condition 3.1 is non-restrictive in the sense that it is necessary so that the candidate weak limit of $\mathbb{C}_{n,C^{(h)}}$ exists pointwise and has continuous sample paths. In the sequel, following Bücher and Volgushev (2013), for any $j \in \{1,\dots,h\}$, we define $\dot{C}^{(h)}_j$ to be zero on the set $\{\boldsymbol{u} \in [0,1]^h : u_j \in \{0,1\}\}$. In addition, as we continue, for any $j \in \{1,\dots,h\}$ and any $\boldsymbol{u} \in [0,1]^h$, $\boldsymbol{u}^{(j)}$ will stand for the vector of $[0,1]^h$ defined by $u^{(j)}_i = u_j$ if $i = j$ and 1 otherwise.

The null weak limit of $\mathbb{C}_{n,C^{(h)}}$ follows in turn from that of the sequential serial empirical process

$$\mathbb{B}_{n,C^{(h)}}(s,t,\boldsymbol{u}) = \frac{1}{\sqrt{n}}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor}\left[\prod_{j=1}^{h}\mathbf{1}\{G(X_{i+j-1}) \le u_j\} - C^{(h)}(\boldsymbol{u})\right], \qquad (s,t,\boldsymbol{u}) \in \Delta\times[0,1]^h, \tag{3.7}$$

with the convention that $\mathbb{B}_{n,C^{(h)}}(s,t,\cdot) = 0$ if $\lfloor nt\rfloor - \lfloor ns\rfloor = 0$.

The following result, stating the weak limit of $\mathbb{C}_{n,C^{(h)}}$ and proven in Appendix B, is a consequence of the results of Bücher and Kojadinovic (2016b) and Bücher *et al.* (2014). It considers $X_1,\dots,X_{n+h-1}$ as a stretch from a *strongly mixing sequence*. For a sequence of random variables $(Z_i)_{i\in\mathbb{Z}}$, the σ-field generated by $(Z_i)_{a\le i\le b}$, $a,b \in \mathbb{Z}\cup\{-\infty,+\infty\}$ is denoted by $\mathcal{F}^b_a$. The strong mixing coefficients corresponding to the sequence $(Z_i)_{i\in\mathbb{Z}}$ are then defined by $\alpha^Z_0 = 1/2$,

$$\alpha^Z_r = \sup_{p\in\mathbb{Z}}\ \sup_{A\in\mathcal{F}^p_{-\infty},B\in\mathcal{F}^{+\infty}_{p+r}}\left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|, \quad r\in\mathbb{N},\ r>0. \tag{3.8}$$

The sequence $(Z_i)_{i\in\mathbb{Z}}$ is said to be *strongly mixing* if $\alpha^Z_r \to 0$ as $r\to\infty$.

Proposition 3.2. Let $X_1,\dots,X_{n+h-1}$ be drawn from a strictly stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of continuous random variables whose strong mixing coefficients satisfy $\alpha^X_r = O(r^{-a})$ for some $a > 1$ as $r\to\infty$. Then, provided Condition 3.1 holds,

$$\sup_{(s,t,\boldsymbol{u})\in\Delta\times[0,1]^h}\left|\mathbb{C}_{n,C^{(h)}}(s,t,\boldsymbol{u}) - \mathbb{B}_{n,C^{(h)}}(s,t,\boldsymbol{u}) + \sum_{j=1}^{h}\dot{C}^{(h)}_j(\boldsymbol{u})\,\mathbb{B}_{n,C^{(h)}}(s,t,\boldsymbol{u}^{(j)})\right| \overset{\mathbb{P}}{\to} 0.$$

Consequently, $\mathbb{C}_{n,C^{(h)}} \rightsquigarrow \mathbb{C}_{C^{(h)}}$ in $\ell^\infty(\Delta\times[0,1]^h)$, where, for any $(s,t,\boldsymbol{u})\in\Delta\times[0,1]^h$,

$$\mathbb{C}_{C^{(h)}}(s,t,\boldsymbol{u}) = \mathbb{B}_{C^{(h)}}(s,t,\boldsymbol{u}) - \sum_{j=1}^{h}\dot{C}^{(h)}_j(\boldsymbol{u})\,\mathbb{B}_{C^{(h)}}(s,t,\boldsymbol{u}^{(j)}), \tag{3.9}$$

and $\mathbb{B}_{C^{(h)}}$ in $\ell^\infty(\Delta\times[0,1]^h)$, a tight centered Gaussian process, is the weak limit of $\mathbb{B}_{n,C^{(h)}}$ in (3.7).

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

Since they are not necessary for the subsequent derivations, the expressions of the covariances of $\mathbb{B}_{C^{(h)}}$ and $\mathbb{C}_{C^{(h)}}$ are not provided. The latter can, however, be deduced from the abovementioned references.

The next result, proven in Appendix B and partly a simple consequence of the previous proposition and the continuous mapping theorem, gives the limiting distribution of $S_{n,C^{(h)}}$ under the null hypothesis of stationarity.

Proposition 3.3. Under the conditions of Proposition 3.2, $\mathbb{D}_{n,C^{(h)}} \rightsquigarrow \mathbb{D}_{C^{(h)}}$ in $\ell^\infty([0,1]^{h+1})$, where, for any $(s,\boldsymbol{u}) \in [0,1]^{h+1}$,

$$\mathbb{D}_{C^{(h)}}(s,\boldsymbol{u}) = \mathbb{C}_{C^{(h)}}(0,s,\boldsymbol{u}) - s\,\mathbb{C}_{C^{(h)}}(0,1,\boldsymbol{u}), \tag{3.10}$$

and $\mathbb{C}_{C^{(h)}}$ is defined by (3.9). As a consequence, we obtain

$$S_{n,C^{(h)}} \rightsquigarrow S_{C^{(h)}} = \sup_{s\in[0,1]} \int_{[0,1]^h} \{\mathbb{D}_{C^{(h)}}(s,\boldsymbol{u})\}^2 \,\mathrm{d}C^{(h)}(\boldsymbol{u}). \tag{3.11}$$

Moreover, the distribution of $S_{C^{(h)}}$ is absolutely continuous with respect to the Lebesgue measure.

3.1.3. Bootstrap and computation of approximate p-values

The null weak limit of $S_{n,C^{(h)}}$ in (3.11) is unfortunately untractable. Starting from Proposition 3.2 and adapting the approach of Bücher and Kojadinovic (2016b) and Bücher *et al.* (2014), we propose to base the computation of approximate p-values for $S_{n,C^{(h)}}$ on *multiplier* resampling versions of $\mathbb{C}_{n,C^{(h)}}$ in (3.6). For any $m \in \mathbb{N}$ and any $(s,t,\boldsymbol{u}) \in \Delta \times [0,1]^h$, let

$$\hat{\mathbb{C}}^{[m]}_{n,C^{(h)}}(s,t,\boldsymbol{u}) = \hat{\mathbb{B}}^{[m]}_{n,C^{(h)}}(s,t,\boldsymbol{u}) - \sum_{j=1}^{h} \dot{C}^{(h)}_{j,1:n}(\boldsymbol{u})\,\hat{\mathbb{B}}^{[m]}_{n,C^{(h)}}(s,t,\boldsymbol{u}^{(j)}), \tag{3.12}$$

where

$$\dot{C}^{(h)}_{j,1:n}(\boldsymbol{u}) = \frac{C^{(h)}_{1:n}(\boldsymbol{u}+h\boldsymbol{e}_j) - C^{(h)}_{1:n}(\boldsymbol{u}-h\boldsymbol{e}_j)}{\min(u_j+h,1) - \max(u_j-h,0)}$$

with $\boldsymbol{e}_j$ the jth unit vector and

$$\hat{\mathbb{B}}^{[m]}_{n,C^{(h)}}(s,t,\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} \xi^{[m]}_{i,n} \left[\prod_{j=1}^{h} \mathbf{1}\{G_{1:n}(X_{i+j-1}) \le u_j\} - C^{(h)}_{1:n}(\boldsymbol{u}) \right], \tag{3.13}$$

with $C^{(h)}_{1:n}$ and $G_{1:n}$ defined by (3.1) and (3.2) respectively. The sequences of random variables $(\xi^{[m]}_{i,n})_{i\in\mathbb{Z}}$, $m \in \mathbb{N}$ appearing in the expressions of the processes $\hat{\mathbb{B}}^{(h),[m]}_n$ in (3.13), $m \in \mathbb{N}$, are independent copies of what was called a *dependent multiplier sequence* in Bücher and Kojadinovic (2016b). Details on that definition, on how such a sequence can be generated and on how a respective block length parameter can be chosen adaptively are presented in Appendix A.

Next, starting from (3.12) and keeping (3.5) in mind, multiplier resampling versions of $\mathbb{D}_{n,C^{(h)}}$ are naturally given, for any $m \in \mathbb{N}$ and $(s,\boldsymbol{u}) \in [0,1]^{h+1}$, by

$$\begin{aligned}\hat{\mathbb{D}}^{[m]}_{n,C^{(h)}}(s,\boldsymbol{u}) &= \lambda_n(s,1)\,\hat{\mathbb{C}}^{[m]}_{n,C^{(h)}}(0,s,\boldsymbol{u}) - \lambda_n(0,s)\,\hat{\mathbb{C}}^{[m]}_{n,C^{(h)}}(s,1,\boldsymbol{u}) \\ &= \hat{\mathbb{C}}^{[m]}_{n,C^{(h)}}(0,s,\boldsymbol{u}) - \lambda_n(0,s)\,\hat{\mathbb{C}}^{[m]}_{n,C^{(h)}}(0,1,\boldsymbol{u}).\end{aligned}$$

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

Corresponding multiplier resampling versions of the statistic $S_{n,C^{(h)}}$ in (3.3) are finally

$$\hat{S}^{[m]}_{n,C^{(h)}} = \sup_{s\in[0,1]} \int_{[0,1]^h} \left\{ \hat{\mathbb{D}}^{[m]}_{n,C^{(h)}}(s,\boldsymbol{u}) \right\}^2 \mathrm{d}C^{(h)}_{1:n}(\boldsymbol{u}), \tag{3.14}$$

which suggests computing an approximate p-value for $S_{n,C^{(h)}}$ as $M^{-1}\sum_{m=1}^{M} \mathbf{1}\big(\hat{S}^{[m]}_{n,C^{(h)}} \geq S_{n,C^{(h)}}\big)$ for some large integer M.

The following proposition establishes the asymptotic validity of the multiplier resampling scheme under the null hypothesis of stationarity. The proof is given in Appendix B.

Proposition 3.4. Assume that $X_1,\ldots,X_{n+h-1}$ are drawn from a strictly stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of continuous random variables whose strong mixing coefficients satisfy $\alpha^X_r = O(r^{-a})$ as $r\to\infty$ for some $a > 3+3h/2$, and $(\xi^{[1]}_{i,n})_{i\in\mathbb{Z}}, (\xi^{[2]}_{i,n})_{i\in\mathbb{Z}},\ldots$ are independent copies of a dependent multiplier sequence satisfying $(\mathcal{M}1)$–$(\mathcal{M}3)$ in Appendix A with $\ell_n = O(n^{1/2-\gamma})$ for some $0<\gamma<1/2$. Then, for any $M\in\mathbb{N}$,

$$\left(\mathbb{C}_{n,C^{(h)}}, \hat{\mathbb{C}}^{[1]}_{n,C^{(h)}},\ldots,\hat{\mathbb{C}}^{[M]}_{n,C^{(h)}}\right) \rightsquigarrow \left(\mathbb{C}_{C^{(h)}}, \mathbb{C}^{[1]}_{C^{(h)}},\ldots,\mathbb{C}^{[M]}_{C^{(h)}}\right)$$

in $\{\ell^\infty(\Delta\times[0,1]^h)\}^{M+1}$, where $\mathbb{C}_{C^{(h)}}$ is defined by (3.9), and $\mathbb{C}^{[1]}_{C^{(h)}},\ldots,\mathbb{C}^{[M]}_{C^{(h)}}$ are independent copies of $\mathbb{C}_{C^{(h)}}$. As a consequence, for any $M\in\mathbb{N}$,

$$\left(\mathbb{D}_{n,C^{(h)}}, \hat{\mathbb{D}}^{[1]}_{n,C^{(h)}},\ldots,\hat{\mathbb{D}}^{[M]}_{n,C^{(h)}}\right) \rightsquigarrow \left(\mathbb{D}_{C^{(h)}}, \mathbb{D}^{[1]}_{C^{(h)}},\ldots,\mathbb{D}^{[M]}_{C^{(h)}}\right)$$

in $\{\ell^\infty([0,1]^{h+1})\}^{M+1}$, where $\mathbb{D}_{C^{(h)}}$ is defined by (3.10), and $\mathbb{D}^{[1]}_{C^{(h)}},\ldots,\mathbb{D}^{[M]}_{C^{(h)}}$ are independent copies of $\mathbb{D}_{C^{(h)}}$. Finally, for any $M\in\mathbb{N}$,

$$\left(S_{n,C^{(h)}}, \hat{S}^{[1]}_{n,C^{(h)}},\ldots,\hat{S}^{[M]}_{n,C^{(h)}}\right) \rightsquigarrow \left(S_{C^{(h)}}, S^{[1]}_{C^{(h)}},\ldots,S^{[M]}_{C^{(h)}}\right),$$

where $S_{C^{(h)}}$ is defined by (3.11), and $S^{[1]}_{C^{(h)}},\ldots,S^{[M]}_{C^{(h)}}$ are independent copies of $S_{C^{(h)}}$.

Notice that, by Lemma 2.2 of Bücher and Kojadinovic (2018) and the continuity of the d.f. of $S_{C^{(h)}}$ (see Proposition 3.3 above), the last statement of Proposition 3.4 is equivalent to the following, more classical, formulation of bootstrap consistency:

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\hat{S}^{[1]}_{n,C^{(h)}} \leq x \mid \boldsymbol{X}_n\right) - \mathbb{P}(S_{n,C^{(h)}} \leq x) \right| \overset{\mathbb{P}}{\to} 0.$$

Furthermore, Lemma 4.2 in Bücher and Kojadinovic (2018) ensures that the test based on $S_{n,C^{(h)}}$ with approximate p-value $p_{n,M}(S_{n,C^{(h)}}) = M^{-1}\sum_{m=1}^{M}\mathbf{1}\big(\hat{S}^{[m]}_{n,C^{(h)}} \geq S_{n,C^{(h)}}\big)$ holds its level asymptotically under the null hypothesis of stationarity as n and M tend to the infinity. By Corollary 4.3 in the same reference, this implies that $p_{n,M_n}(S_{n,C^{(h)}}) \rightsquigarrow$ Uniform(0, 1) when $n\to\infty$ for any sequence $M_n\to\infty$.

3.2. A d.f.-based test sensitive to changes in the contemporary distribution

We propose to combine the previous test with a test particularly sensitive to departures from $H^{(1)}_0$ in (1.3). As mentioned in the introduction, a natural candidate is the CUSUM test studied in Gombay and Horváth (1999) and extended in Holmes *et al.* (2013). For the sake of a simpler presentation, we proceed as if the only available observations were $X_1,\ldots,X_n$, thereby ignoring the remaining $h-1$ ones. The test statistic can then be written as

$$S_{n,G} = \sup_{s\in[0,1]} \int_{\mathbb{R}} \left\{\mathbb{E}_n(s,x)\right\}^2 \mathrm{d}G_{1:n}(x), \tag{3.15}$$

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

where

$$\mathbb{E}_n(s,x) = \sqrt{n}\lambda_n(0,s)\lambda_n(s,1)\left\{G_{1:\lfloor ns\rfloor}(x) - G_{\lfloor ns\rfloor+1:n}(x)\right\}, \qquad (s,x) \in [0,1]\times\mathbb{R}, \tag{3.16}$$

and, for any $1 \le k \le l \le n$, $G_{k:l}$ is defined as in (3.2) but with $h = 1$. As one can see, the test involves the comparison of the empirical d.f. of $X_1, \dots, X_k$ with the one of $X_{k+1}, \dots, X_n$ for all $k \in \{1, \dots, n-1\}$. Under $H_0^{(1)}$ in (1.3), it can be verified that $\mathbb{E}_n$ in (3.16) can be written as

$$\mathbb{E}_n(s,x) = \mathbb{G}_n(s,x) - \lambda_n(0,s)\,\mathbb{G}_n(1,x), \qquad (s,x) \in [0,1]\times\mathbb{R},$$

where

$$\mathbb{G}_n(s,x) = \sqrt{n}\,\lambda_n(0,s)\,\{G_{1:\lfloor ns\rfloor}(x) - G(x)\}, \qquad (s,x) \in [0,1]\times\mathbb{R}. \tag{3.17}$$

The following result, proven in Appendix B and providing the null weak limit of $S_{n,G}$ in (3.15), is partly an immediate consequence of Theorem 1 of Bücher (2015) and of the continuous mapping theorem.

Proposition 3.5. Let $X_1, \dots, X_n$ be drawn from a strictly stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of continuous random variables whose strong mixing coefficients satisfy $\alpha_r = O(r^{-a})$ for some $a > 1$, as $r \to \infty$. Then, $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ in $\ell^\infty([0,1] \times\mathbb{R})$, where $\mathbb{G}$ is a tight centered Gaussian process with covariance function

$$\mathrm{Cov}\{\mathbb{G}(s,x), \mathbb{G}(t,y)\} = \min(s,t)\sum_{k\in\mathbb{Z}} \mathrm{Cov}\{\mathbf{1}(X_0 \le x)\mathbf{1}(X_k \le y)\}.$$

Consequently, $\mathbb{E}_n \rightsquigarrow \mathbb{E}$ in $\ell^\infty([0,1]\times\mathbb{R})$, where

$$\mathbb{E}(s,x) = \mathbb{G}(s,x) - s\mathbb{G}(1,x), \qquad (s,x) \in [0,1]\times\mathbb{R}, \tag{3.18}$$

and $S_{n,G} \rightsquigarrow S_G$ with

$$S_G = \sup_{s\in[0,1]} \int_{\mathbb{R}} \{\mathbb{E}(s,x)\}^2\,\mathrm{d}G(x). \tag{3.19}$$

Moreover, the distribution of S_G is absolutely continuous with respect to the Lebesgue measure.

Following Gombay and Horváth (1999), Holmes *et al.* (2013) and Bücher and Kojadinovic (2016b), we shall compute approximate p-values for $S_{n,G}$ using multiplier resampling versions of $\mathbb{G}_n$ in (3.17). Let $(\xi_{i,n}^{[m]})_{i\in\mathbb{Z}}$, $m \in \mathbb{N}$ be independent copies of the same dependent multiplier sequence. For any $m \in \mathbb{N}$ and any $(s,x) \in [0,1]\times\mathbb{R}$, let

$$\begin{aligned}
\hat{\mathbb{G}}_n^{[m]}(s,x) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} \xi_{i,n}^{[m]}\left\{\mathbf{1}(X_i \le x) - G_{1:n}(x)\right\}, \\
\hat{\mathbb{E}}_n^{[m]}(s,x) &= \mathbb{G}_n^{[m]}(s,x) - \lambda_n(0,s)\,\mathbb{G}_n^{[m]}(1,x), \\
\hat{\mathbb{S}}_{n,G}^{[m]} &= \sup_{s\in[0,1]}\int_{\mathbb{R}} \left\{\hat{\mathbb{E}}_n^{[m]}(s,x)\right\}^2 \mathrm{d}G_{1:n}(x).
\end{aligned} \tag{3.20}$$

An approximate p-value for $S_{n,G}$ will then be computed as $p_{n,M}(S_{n,G}) = M^{-1}\sum_{m=1}^M \mathbf{1}\left(\hat{S}_{n,G}^{[m]} \ge S_{n,G}\right)$ for some large integer M. The asymptotic validity of this approach under the null hypothesis of stationarity can be shown as for the test based on $S_{n,C^{(h)}}$ presented in the previous section. The result is a direct consequence of Corollary 2.2 in Bücher and Kojadinovic (2016b); see also Proposition 3.6 in the next section. In particular, $p_{n,M_n}(S_{n,G}) \rightsquigarrow \mathrm{Uniform}(0,1)$ when $n \to \infty$, for any sequence $M_n \to \infty$.

DOI: 10.1111/jtsa.12431
© 2018 John Wiley & Sons Ltd

3.3. Combining the two tests

To combine the two tests, we use the general procedure described in Section 2 with $r = 2$, $T_{n,1} = S_{n,C^{(h)}}$ and $T_{n,2} = S_{n,G}$, for some suitable function $\psi : (0,1)^2 \to \mathbb{R}$ such as ψ_S in (2.1) or ψ_F in (2.2). To be able to apply Proposition 2.1, we need to find conditions under which $\boldsymbol{T}_n = (T_{n,1}, T_{n,2})$ and its bootstrap replicates satisfy (2.7) or, equivalently, (2.8). A natural prerequisite is to compute the M bootstrap replicates of $T_{n,1} = S_{n,C^{(h)}}$ and $T_{n,2} = S_{n,G}$ in (3.14) and (3.20) respectively, using the same M dependent multiplier sequences. Since a moving average approach is used to generate such sequences, it follows from (A.1) that it is sufficient to impose that the same M initial independent normal sequences be used for both tests. In practice, prior to using (A.1) to generate the M independent copies of the same dependent multiplier sequence, we estimate the key bandwidth parameter ℓ_n from $X_1, \ldots, X_{n+h-1}$ using the approach proposed in Bücher and Kojadinovic (2016b, Section 5.1), briefly overviewed in Appendix A.

The next result, proven in Appendix B, provides conditions under which (2.7) holds.

Proposition 3.6. Under the conditions of Proposition 3.4, for any $M \in \mathbb{N}$,

$$\left((\mathbb{D}_{n,C^{(h)}}, \mathbb{E}_n), (\hat{\mathbb{D}}^{[1]}_{n,C^{(h)}}, \hat{\mathbb{E}}^{[1]}_n), \ldots, (\hat{\mathbb{D}}^{[M]}_{n,C^{(h)}}, \hat{\mathbb{E}}^{[M]}_n)\right) \rightsquigarrow \left((\mathbb{D}_{C^{(h)}}, \mathbb{E}), (\mathbb{D}^{[1]}_{C^{(h)}}, \mathbb{E}^{[1]}), \ldots, (\mathbb{D}^{[M]}_{C^{(h)}}, \mathbb{E}^{[M]})\right)$$

in $\{\ell^\infty([0,1] \times \mathbb{R})\}^{2(M+1)}$, where $\mathbb{D}_{C^{(h)}}$ and $\mathbb{E}$ are defined by (3.10) and (3.18) respectively, and $(\mathbb{D}^{[1]}_{C^{(h)}}, \mathbb{E}^{[1]}), \ldots, (\mathbb{D}^{[M]}_{C^{(h)}}, \mathbb{E}^{[M]})$ are independent copies of $(\mathbb{D}_{C^{(h)}}, \mathbb{E})$. Note that we do not specify the joint law of $(\mathbb{D}_{C^{(h)}}, \mathbb{E})$; it will only be important that $(\hat{\mathbb{D}}^{[m]}_{n,C^{(h)}}, \hat{\mathbb{E}}^{[m]}_n)$, $m \in \{1, \ldots, M\}$ can be considered to have the same joint law as $(\mathbb{D}_{C^{(h)}}, \mathbb{E})$ asymptotically. As a consequence,

$$\left((S_{n,C^{(h)}}, S_{n,G}), (\hat{S}^{[1]}_{n,C^{(h)}}, \hat{S}^{[1]}_{n,G}), \ldots, (\hat{S}^{[M]}_{n,C^{(h)}}, \hat{S}^{[M]}_{n,G})\right) \rightsquigarrow \left((S_{C^{(h)}}, S_G), (S^{[1]}_{C^{(h)}}, S^{[1]}_G), \ldots, (S^{[M]}_{C^{(h)}}, S^{[M]}_G)\right),$$

where $S_{C^{(h)}}$ and S_G are defined by (3.11) and (3.19) respectively, and where the random vectors $(S^{[1]}_{C^{(h)}}, S^{[1]}_G), \ldots, (S^{[M]}_{C^{(h)}}, S^{[M]}_G)$ are independent copies of $(S_{C^{(h)}}, S_G)$.

A consequence of the previous proposition is that the unconditional bootstrap consistency statement in (2.9) holds under the conditions of Proposition 3.4. To conclude that the conditional statements given in (2.11) and (2.12) hold as well, it is necessary to establish that W, given generically by (2.10), has a continuous d.f. Proving the latter might actually be quite complicated: unlike $S_{C^{(h)}}$ in (3.11) and S_G in (3.19), W is not a convex function of some Gaussian process, whence the general results from Davydov and Lifshits (1984) and the references therein are not applicable. Proving the absolute continuity of the vector $(S_{C^{(h)}}, S_G)$ could be a first step, but the latter does not seem easy either: available results in the literature are mostly based on complicated conditions from Malliavin Calculus (see, e.g., Theorem 2.1.2 in Nualart, 2006). For these reasons, we do not pursue such investigations any further in this paper. Nonetheless, we conjecture that W will have a continuous d.f. in all except a few very pathological situations.

Under suitable conditions on alternative models, it can further be shown that at least one of the statistics $S_{n,G}$ or $S_{n,C^{(h)}}$ (for h suitably chosen) diverges to infinity in probability at rate n. For instance, for $S_{n,G}$, under the assumption of at most one change in the contemporary d.f. of the time series, the latter can be shown by adapting to the serially dependent case the arguments used in Holmes *et al.* (2013, Proof of Theorem 3(i)). Further details are omitted for the sake of brevity. As far as bootstrap replicates of $S_{n,G}$ or $S_{n,C^{(h)}}$ are concerned, based on our extensive simulation results, we conjecture that, for many alternative models, the bootstrap replicates are of lower order than $O_{\mathbb{P}}(n)$. As a consequence, assuming the aforementioned results, and when the combining function ψ is ψ_F in (2.2), one can rely on Proposition 2.2 to show the consistency of the test based on $W^{[0]}_{n,M_n}$ in (2.4).

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

3.4. On the choice of the embedding dimension h

The methodology described in the previous sections depends on the embedding dimension h. In this section, we will provide some intuition about the trade-off between the choice of small and large values of h. Based on the developed arguments, and on the large-scale simulation study in Section 5 and in Appendix S1, we will make a practical suggestion at the end of this section.

Let us start by considering arguments in favor of choosing a large value of h. For that purpose, note that stationarity is equivalent to $H_0^{(1)}$ in (1.3) and $H_{0,c}^{(h)}$ in (1.4) for all $h \geq 2$ and that a test based on the embedding dimension h can only detect alternatives for which $H_{0,c}^{(h)}$ does not hold. Hence, since $H_{0,c}^{(2)} \Leftarrow H_{0,c}^{(3)} \Leftarrow \dots$, we would like to choose h as large as possible to be consistent against as many alternatives as possible. Note that, at the same time, the potential gain in moving from h to $h+1$ should decrease with h: first, the larger h, the less likely it seems that real-life phenomena satisfy $H_{0,c}^{(h)}$ but not $H_{0,c}^{(h+1)}$; second, from a model engineering perspective, the larger the value of h, the more difficult and artificial it becomes to construct sensible models that satisfy $H_{0,c}^{(h)}$ but not $H_{0,c}^{(h+1)}$. To illustrate the latter point, constructing such a model on the level of copulas would amount to finding (at least two) different $(h+1)$-dimensional copulas $C^{(h+1)}$ that have the same lower-dimensional (multi-variate) margins. More formally and given the serial context under consideration, this would mean finding a model such that

$$C^{(h+1)}(1, \dots, 1, u_i, \dots, u_{i+k-1}, 1, \dots, 1) = C^{(k)}(u_i, \dots, u_{i+k-1}),$$

for all $k \in \{2, \dots, h\}$, $i \in \{1, \dots, h-k+2\}$, $u_i, \dots, u_{i+k-1} \in [0,1]$ for some given k-dimensional copulas $C^{(k)}$. This problem is closely related to the so-called compatibility problem (Nelsen, 2006, Section 3.5), and to the best of our knowledge, a general solution has not yet been found. Some necessary conditions can be found in Rüschendorf (1985, Theorem 4) for the case of copulas that are absolutely continuous with respect to the Lebesgue measure on the unit hypercube. As another example, consider as a starting point the autoregressive process $X_i = \beta X_{i-h} + \varepsilon_i$, where the noises $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$ are i.i.d. and where $|\beta| < 1$. The components of the vectors $\boldsymbol{Y}_i^{(h)} = (X_i, \dots, X_{i+h-1})$ are then i.i.d. $\mathcal{N}(0, \tau^2/(1-\beta^2))$. Hence, $C^{(h)}$ is the independence copula, and $H_0^{(h)}$ in (1.2) is met, while $H_{0,c}^{(h+1)}$ in (1.4) would not be met should the parameters τ and β change (smoothly or abruptly) in such a way that $\tau^2/(1-\beta^2)$ stays constant, a rather artificial example. More generally, one could argue that, the larger h, the more artificial instances of common time series models (such as ARMA- or GARCH-type models) for which $H_{0,c}^{(h)}$ holds but not $H_{0,c}^{(h+1)}$.

The previous paragraph suggests choosing h as large as possible, even if the marginal gain of an increase of h becomes smaller for larger and larger h. On the contrary, there are also good reasons for choosing rather small h. Indeed, for many sensible models, the power of the test based on $S_{n,C^{(h)}}$ in (3.3) is a decreasing function of h, at least from some small value onward. This observation will, for instance, be one of the results of our simulation study in Section 5 (see, e.g., Figure 1), but it can also be supported by more theoretical arguments. Indeed, for instance, consider the following simple alternative model: $X_1, X_2, \dots$ have the same d.f. G and, for some $s^* \in (0,1)$, $\boldsymbol{Y}_i^{(h)}$, $i \in \{1, \dots, \lfloor ns^* \rfloor - \lfloor h/2 \rfloor\}$, have copula $C_1^{(h)}$ and $\boldsymbol{Y}_i^{(h)}$, $i \in \{\lfloor ns^* \rfloor + 1 + \lfloor h/2 \rfloor, \dots, n\}$ have copula $C_2^{(h)} \neq C_1^{(h)}$. For simplicity, we do not specify the laws of the $\boldsymbol{Y}_i^{(h)}$ for $i \in \{\lfloor ns^* \rfloor - \lfloor h/2 \rfloor + 1, \dots, \lfloor ns^* \rfloor + \lfloor h/2 \rfloor\}$ (these observations induce negligible effects in the following reasoning), whence, asymptotically, we can do 'as if' $\boldsymbol{Y}_i^{(h)}$, $i \in \{1, \dots, \lfloor ns^* \rfloor\}$, have copula $C_1^{(h)}$ and $\boldsymbol{Y}_i^{(h)}$, $i \in \{\lfloor ns^* \rfloor + 1, \dots, n\}$, have copula $C_2^{(h)}$. Under this model and additional regularity conditions, we observe that

$$n^{-1} S_{n,C^{(h)}} \rightsquigarrow \kappa_h \equiv \{s^*(1-s^*)\}^2 \int_{[0,1]^h} \{C_1^{(h)}(\boldsymbol{u}) - C_2^{(h)}(\boldsymbol{u})\}^2 \, \mathrm{d}C_{s^*}^{(h)}(\boldsymbol{u}),$$

where $C_{s^*}^{(h)} = s^* C_1^{(h)} + (1-s^*) C_2^{(h)}$. In other words, the dominating term in an asymptotic expansion of $S_{n,C^{(h)}}$ diverges to infinity at rate n, with scaling factor κ_h depending on h. Since we conjecture that the bootstrap replicates of $S_{n,C^{(h)}}$ are of a lower order than $O_{\mathbb{P}}(n)$ for any h, we further conjecture that the power curves of the test will be controlled to a large extent by the 'signal of non-stationarity' κ_h. The impact of h on this quantity is ambiguous, but in many sensible models, it is decreasing in h eventually, inducing a sort of 'curse of dimensionality'. This

DOI: 10.1111/jtsa.12431
© 2018 John Wiley & Sons Ltd

results in a smaller power of the corresponding test for larger h and fixed sample size n, as will be empirically confirmed in several scenarios considered in the Monte Carlo experiments of Section 5 and in Appendix S1.

In addition, several arguments lead us to assume that smaller values of h also yield a better approximation of the nominal level. From an empirical perspective, this will be confirmed for all the scenarios under stationarity in our Monte Carlo experiments. While we are not aware of any theoretical result for our quite general serially dependent setting (that would include the dependent multiplier bootstrap), some results are available for the i.i.d. or non-bootstrap case. For instance, Chernozhukov *et al.* (2013) provide bounds on the approximation error of i.i.d. sum statistics by an i.i.d. multiplier bootstrap; the bounds are increasing in the dimension h. Moreover, with the asymptotics of our test statistics relying on the asymptotics of empirical processes, we would be interested in a good approximation of empirical processes by their limiting counterparts. As shown in Dedecker *et al.* (2014) for the case of beta-mixing random variables, the approximation error of strong approximation techniques is again increasing in h.

Globally, the above arguments, as well as the results of the simulation study in Section 5 below and in Appendix S1, suggest that a rather small value of h, for instance, in $\{2,3,4\}$, should be sufficient to test strong stationarity in many situations. Such a choice would provide relatively powerful tests for many interesting alternatives without strongly suffering from the curse of dimensionality. Depending on the ultimate interest, one might also consider choosing h differently, for example, as the 'forecast horizon'. Finally, a natural research direction would consist of developing data-driven procedures for choosing h, for instance, following ideas developed in Escanciano and Lobato (2009) for testing serial correlation in a time series. However, such an analysis appears to be a research topic in itself and lies beyond the scope of the present paper.

4. A COMBINED TEST OF SECOND-ORDER STATIONARITY

Starting from the general framework considered in Bücher and Kojadinovic (2016a) and proceeding as in Section 3, one can derive a combined test of second-order stationarity. Given the embedding dimension $h \geq 2$ and the available univariate observations $X_1, \dots, X_{n+h-1}$, let $\boldsymbol{Z}_i^{(q)}$, $i \in \{1, \dots, n\}$ be the random variables defined by

$$\boldsymbol{Z}_i^{(q)} = \begin{cases} X_i, & \text{if } q = 1, \\ (X_i, X_{i+q-1}), & \text{if } q \in \{2, \dots, h\}. \end{cases} \tag{4.1}$$

Let ϕ be a symmetric, measurable function on $\mathbb{R} \times \mathbb{R}$ or on $\mathbb{R}^2 \times \mathbb{R}^2$. Then, the U-statistic of order 2 with kernel ϕ obtained from the subsample $\boldsymbol{Z}_k^{(q)}, \dots, \boldsymbol{Z}_l^{(q)}$, $1 \leq k < l \leq n$ is given by

$$U_{k:l,q,\phi,} = \frac{1}{\binom{l-k+1}{2}} \sum_{k \leq i < j \leq l} \phi\big(\boldsymbol{Z}_i^{(q)}, \boldsymbol{Z}_j^{(q)}\big). \tag{4.2}$$

We focus on CUSUM tests for change-point detection based on the generic statistic

$$S_{n,q,\phi} = \max_{2 \leq k \leq n-2} |\mathbb{U}_{n,q,\phi}(k/n)| = \sup_{s \in [0,1]} |\mathbb{U}_{n,q,\phi}(s)|, \tag{4.3}$$

where

$$\mathbb{U}_{n,q,\phi}(s) = \sqrt{n}\lambda_n(0,s)\lambda_n(s,1)(U_{1:\lfloor ns \rfloor,q,\phi} - U_{\lfloor ns \rfloor+1:n,q,\phi}) \quad \text{if } s \in [2/n, 1-2/n],$$

and $\mathbb{U}_{n,q,\phi}(s) = 0$ otherwise.

With the aim of assessing whether second-order stationarity is plausible, the following possibilities for $q \in \{1, \dots, h\}$ and the kernel ϕ are of interest: if $q = 1$ and $\phi(z, z') = m(z, z') = z, z' \in \mathbb{R}$, the statistic $S_{n,q,\phi} = S_{n,1,m}$ is (asymptotically equivalent to) the classical CUSUM statistic that is particularly sensitive to changes in the

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

expectation of $X_1, \dots, X_n$. Similarly, setting $q = 1$ and $\phi(z, z') = v(z, z') = (z - z')^2/2$, $z, z' \in \mathbb{R}$ gives rise to the statistic $S_{n,1,v}$, particularly sensitive to changes in the variance of the observations. For $q \in \{2, \dots, h\}$, setting $\phi(z, z') = a(z, z') = (z_1 - z_1')(z_2 - z_2')/2$, $z, z' \in \mathbb{R}^2$ results in the CUSUM statistic $S_{n,q,a}$, sensitive to changes in the autocovariance at lag $q - 1$.

From Bücher and Kojadinovic (2016a), CUSUM tests based on $S_{n,1,m}$, $S_{n,1,v}$ and $S_{n,q,a}$, $q \in \{2, \dots, h\}$, sensitive to changes in the expectation, variance and autocovariances respectively, can all be carried out using a resampling scheme based on dependent multiplier sequences. As a consequence, they can be combined by proceeding as in Sections 2 and 3.3. Specifically, for the generic test based on $S_{n,q,\phi}$, let $(\xi_{i,n}^{[m]})_{i \in \mathbb{Z}}$, $m \in \mathbb{N}$ be independent copies of the same dependent multiplier sequence, and for any $m \in \mathbb{N}$ and $s \in [0, 1]$, let

$$\hat{\mathbb{U}}_{n,q,\phi}^{[m]}(s) = \frac{2}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_{i,n}^{(m)} \hat{\phi}_{1,1:n}\big(\boldsymbol{Z}_i^{(q)}\big) - \lambda_n(0, s) \times \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i,n}^{(m)} \hat{\phi}_{1,1:n}\big(\boldsymbol{Z}_i^{(q)}\big), \quad \text{if } s \in [2/n, 1 - 2/n],$$

and $\hat{\mathbb{U}}_{n,q,\phi}^{[m]}(s) = 0$ otherwise, where

$$\hat{\phi}_{1,1:n}\big(\boldsymbol{Z}_i^{(q)}\big) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} \phi\big(\boldsymbol{Z}_i^{(q)}, \boldsymbol{Z}_j^{(q)}\big) - U_{1:n,q,\phi}, \quad i \in \{1, \dots, n\},$$

with $U_{1:n,q,\phi}$ defined by (4.2). Then, multiplier replications of $S_{n,q,\phi}$ are given by

$$\hat{\mathbb{S}}_{n,q,\phi}^{[m]} = \max_{2 \le k \le n-2} \left| \hat{\mathbb{U}}_{n,q,\phi}^{[m]}(k/n) \right| = \sup_{s \in [0,1]} \left| \hat{\mathbb{U}}_{n,q,\phi}^{[m]}(s) \right|, \qquad m \in \mathbb{N},$$

and an approximate p-value for $S_{n,q,\phi}$ can be computed as $p_{n,M}(S_{n,q,\phi}) = M^{-1} \sum_{m=1}^{M} \mathbf{1}\big(\hat{S}_{n,q,\phi}^{[m]} \ge S_{n,q,\phi}\big)$ for some large integer $M \in \mathbb{N}$.

To obtain a test of second-order stationarity, we again use the combining procedure of Section 2, this time, with $r = h + 1$, $T_{n,1} = S_{n,1,m}$, $T_{n,2} = S_{n,1,v}$ and $T_{n,q+1} = S_{n,q,a}$, $q \in \{2, \dots, h\}$ for some function $\psi : (0, 1)^{h+1} \to \mathbb{R}$ decreasing in each of its arguments such as ψ_S in (2.1) or ψ_F in (2.2). As in Section 3.3, to compute bootstrap replicates of the components of $\boldsymbol{T}_n = (T_{n,1}, \dots, T_{n,r})$, we use the same M dependent multiplier sequences. Specifically, we first estimate ℓ_n from $X_1, \dots, X_n$ as explained in Bücher and Kojadinovic (2016a, Section 2.4) for $\phi = m$. Then, with the obtained value of ℓ_n, we generate M independent copies of the same dependent multiplier sequence using (A.1) and compute the corresponding multiplier replicates $\hat{S}_{n,q,\phi}^{[1]}, \dots, \hat{S}_{n,q,\phi}^{[M]}$ for $q = 1$ and $\phi \in \{m, v\}$, and for $q \in \{2, \dots, h\}$ and $\phi = a$.

As given in Section 3.3, to establish the asymptotic validity of the global test under stationarity using Proposition 2.1, we need to establish conditions under which, using the notation of Section 2, $\boldsymbol{T}_n = (T_{n,1}, \dots, T_{n,r})$ and its bootstrap replicates satisfy (2.7) or, equivalently, (2.8). The latter can be proven by starting from Proposition 2.5 in Bücher and Kojadinovic (2016a) and by proceeding as in the proofs of the results stated in Section 3.3. For the sake of simplicity, the conditions in Proposition 4.1 require that $X_1, \dots, X_{n+h-1}$ is a stretch from an absolutely regular sequence. Indeed, assuming that $(X_i)_{i \in \mathbb{Z}}$ is only strongly mixing leads to significantly more complex statements. Recall that the absolute regularity coefficients corresponding to a sequence $(Z_i)_{i \in \mathbb{Z}}$ are defined by

$$\beta_r^Z = \sup_{p \in \mathbb{Z}} \mathrm{E} \sup_{A \in \mathcal{F}_{p+r}^{\infty}} \left| \mathbb{P}\big(A \mid \mathcal{F}_{-\infty}^{p}\big) - \mathbb{P}(A) \right|, \quad r \in \mathbb{N}, r > 0,$$

where $\mathcal{F}_a^b$ is defined above (3.8). The sequence $(Z_i)_{i \in \mathbb{N}}$ is then said to be *absolutely regular* if $\beta_r \to 0$ as $r \to \infty$. As $\alpha_r^Z \le \beta_r^Z$, absolute regularity implies strong mixing.

DOI: 10.1111/jtsa.12431
© 2018 John Wiley & Sons Ltd

Proposition 4.1. Let $X_1, \dots, X_{n+h-1}$ be drawn from a strictly stationary sequence $(X_i)_{i\in\mathbb{Z}}$ such that $\mathrm{E}\{|X_1|^{2(4+\delta)}\} < \infty$ for some $\delta > 0$. In addition, let $(\xi_{i,n}^{[1]})_{i\in\mathbb{Z}}$ and $(\xi_{i,n}^{[2]})_{i\in\mathbb{Z}}$ be independent copies of a dependent multiplier sequence satisfying $(\mathcal{M}1)$–$(\mathcal{M}3)$ in Appendix A with $\ell_n = O(n^{1/2-\gamma})$ for some $1/(6+2\delta) < \gamma < 1/2$. Then, if $\beta_r^X = O(r^{-b})$ for some $b > 2(4+\delta)/\delta$ as $r \to \infty$, (2.7) or, equivalently, (2.8) hold.

5. MONTE CARLO EXPERIMENTS

Extensive simulations were carried out to try to answer several fundamental questions (hereafter in bold) regarding the tests proposed in Sections 3 and 4. For the sake of readability, we only present a small subset of the performed Monte Carlo experiments in detail and refer the reader to Appendix S1 for more results. Before formulating the questions, we introduce abbreviations for the components tests whose behavior we investigated:

- d for the d.f. test based on $S_{n,G}$ in (3.15);
- c for the empirical autocopula test at lag $h-1$ based on $S_{n,C^{(h)}}$ in (3.3) (the value of h will always be clear from the context);
- m for the sample mean test based on $S_{n,m}^{(1)}$ defined generically by (4.3);
- v for the variance test based on $S_{n,v}^{(1)}$ defined generically by (4.3) and
- a for the autocovariance test at lag $q-1$ based on $S_{n,a}^{(q)}$, $q \in \{2, \dots, h\}$, defined generically by (4.3) (the value of q will always be clear from the context).

With these conventions, the following abbreviations are used for the combined tests:

- dc: equally weighted combination of the tests d and c for embedding dimension h or, equivalently, lag $h-1$;
- va: combination of the test v with weight 1/2 and the autocovariance tests a for lags $q \in \{1, \dots, h-1\}$ with equal weights $1/\{2(h-1)\}$;
- mva: combination of the test m with weight 1/3, of the variance test v with weight 1/3 and the autocovariance tests a for lags $q \in \{1, \dots, h-1\}$ with equal weights $1/\{3(h-1)\}$; and
- dcp: combination of the test d with weight 1/2 with pairwise bivariate empirical autocopula tests for lags $1, \dots, h-1$ with equal weights $1/\{2(h-1)\}$; in other words, the d.f. test based on $S_{n,G}$ in (3.15) is combined with $S_{n,C^{(2)}}$ in (3.3) and $S_{n,\bar{C}^{(3)}}, \dots, S_{n,\bar{C}^{(h)}}$, where the latter are the analogues of $S_{n,C^{(2)}}$ but for lags $2, \dots, h-1$ (i.e., they are computed from (4.1) for $q \in \{3, \dots, h\}$).

The above choices for the weights are arbitrary and thus clearly debatable. An 'optimal' strategy for the choice of the weights is beyond the scope of this work. For the function ψ in Sections 3 and 4, we only consider ψ_F in (2.2) as the use of ψ_S in (2.1) sometimes gave inflated levels.

Let us now state the fundamental questions concerning the studied tests that we attempted to answer empirically by means of a large number of Monte Carlo experiments.

Do the studied component and combined tests maintain their level? As is explained in detail in Appendix S1, 10 strictly stationarity models, including ARMA, GARCH and nonlinear autoregressive models with either normal or Student t with 4 degrees of freedom innovations, were used to generate observations under the null hypothesis of stationarity. The rank-based tests of Section 3, that is, d, c, dc and dcp, were never found to be too liberal, while some of the second-order tests of Section 4, namely, v, va and mva, were found to reject stationarity too often for a particular GARCH model mimicking S&P500 daily log-returns.

How do the rank-based tests of Section 3 compare to the second-order tests of Section 4 in terms of power? As presented in detail in Appendix S1, to investigate the power of the tests, eight models connected to the literature on locally stationary processes were considered along with four models more in line with the change-point detection literature. All tests were found to have reasonable power for at least one (and, usually, several) of the alternatives under consideration. The combined rank-based tests proposed in Section 3, that is, dc or dcp, were found, overall, to be more powerful than the combined second-order tests, namely,

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

Table I. Percentages of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n = 128$ from model D(σ), S(β) or DS(σ, β) for various values of σ and β. The meanings of the abbreviations d, c, dc are given in Section 5

	$h = 2$ or lag 1		
Model	d	c	dc
D(2): 'Small change in contemporary dist. only'	33.6	2.2	16.4
D(3): 'Large change in contemporary dist. only'	81.6	1.6	59.2
S(0.3): 'Small change in serial dep. only'	6.4	19.6	16.6
S(0.9): 'Large change in serial dep. only'	13.8	64.2	62.8
DS(2, 0.4): 'Small change in both'	17.2	28.8	35.4
DS(4, 0.7): 'Large change in both'	75.6	70.0	92.6

va or mva, even in situations involving changes in the second-order characteristics of the underlying time series.

How are the powers of the proposed component and combined tests related? For the sake of illustration, we only focus on the component tests d and c, and the combined test dc, and consider three simple data generating models:

- D(σ) - 'Change in the contemporary distribution only': The $n/2$ first observations are i.i.d. from the N$(0, \sigma^2)$ distribution, and the $n/2$ last observations are i.i.d. from the N$(0, 1)$ distribution.
- S(β) - 'Change in the serial dependence only': The $n/2$ first observations are i.i.d. standard normal, and the $n/2$ last observations are drawn from an AR(1) model with parameter β and centered normal innovations with variance $(1 - \beta^2)$. The contemporary distribution is thus constant and equal to the standard normal.
- DS(σ, β) - 'Change in the contemporary distribution and the serial dependence': The $n/2$ first observations are i.i.d. from the N$(0, \sigma^2)$ distribution, and the $n/2$ last observations are drawn from an AR(1) model with parameter β and N$(0, 1)$ innovations.

At the 5% significance level, the rejection percentages of the null hypothesis of stationarity computed from 1000 samples of size $n = 128$ from model D(σ), S(β) or DS(σ, β) for various values of σ and β are given in Table I for the tests d, c and dc for $h = 2$. As one can see from the first four rows of the table, when one of the component tests has hardly any power, a 'dampening effect' occurs for the combined test. However, when the two components tests tend to detect changes, most of the time, a 'reinforcement effect' seems to simultaneously occur for the combined test as can be seen from the last two rows of the table.

Is the combined test dc truly more powerful than a simple multi-variate extension of the test d designed to be directly sensitive to departures from $H_0^{(h)}$ in (1.2)? Note that, to implement the latter test for a given embedding dimension h, it suffices to proceed as in Section 3.2 but by using the h-dimensional empirical d.f.s of the h-dimensional random vectors $\boldsymbol{Y}_i^{(h)}$ in (1.1) instead of the one-dimensional empirical d.f.s generically given by (3.2). Let dh be the abbreviation of this test. To provide an empirical answer to the above question, we consider a similar setup as previously described. The rejection percentages of the null hypothesis of stationarity computed from 1000 samples of size $n = 128$ from model D(σ), S(β) or DS(σ, β) for various values of σ and β are given in Table II for the tests dc, dcp and dh for $h \in \{2, 3\}$. As one can see, the test dh seems to have hardly any power when the non-stationarity is only due to a change in the serial dependence. Furthermore, even when the non-stationarity results from a change in the contemporary distribution, the test dh appears to be less powerful, overall, than the combined tests dc and dcp.

What is the influence of the choice of the embedding dimension h on the empirical levels and the powers of the proposed tests? The extensive simulations results available in Appendix S1 indicate that, under the null hypothesis of stationarity, tests c and dc tend, overall, to become more and more conservative as h increases for fixed sample size n. For fixed h, the empirical levels seem to get closer to the 5% nominal level as

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

Table II. Percentages of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n = 128$ from model D(σ), S(β) or DS(σ, β) for various values of σ and β. The meanings of the abbreviations dc, dcp and dh are given in Section 5

	$h = 2$		$h = 3$ or lag 2		
Model	dc	dh	dc	dcp	dh
D(2): 'Small change in contemporary dist. only'	16.4	21.8	17.8	26.6	24.8
D(3): 'Large change in contemporary dist. only'	59.2	52.4	58.8	73.0	44.0
S(0.3): 'Small change in serial dep. only'	16.6	7.2	18.2	13.0	9.0
S(0.9): 'Large change in serial dep. only'	62.8	15.6	63.0	65.0	16.0
DS(2, 0.4): 'Small change in both'	35.4	20.6	42.2	34.8	30.0
DS(4, 0.7): 'Large change in both'	92.6	67.6	92.4	91.6	71.6

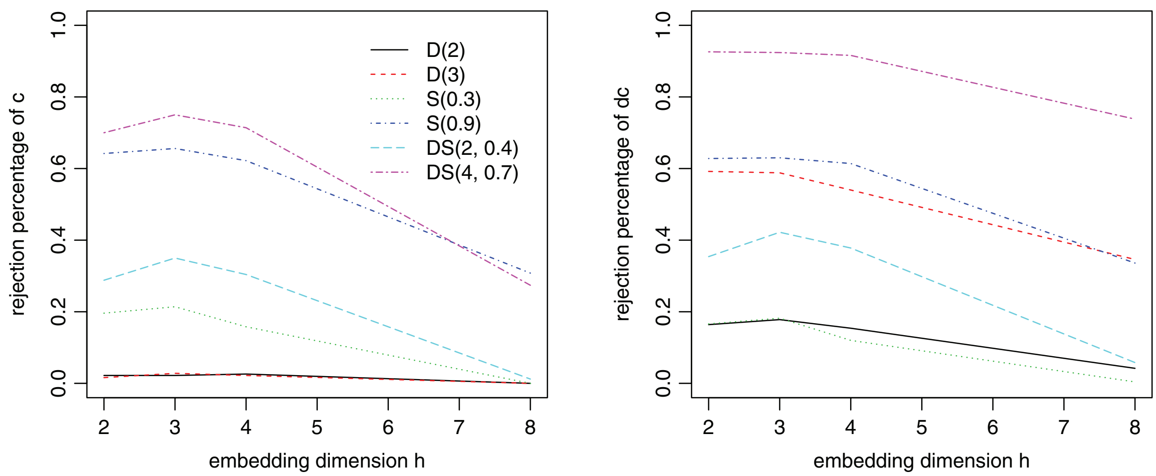

Figure 1. Rejection percentages of c and dc against the embedding dimension $h \in \{2, 3, 4, 8\}$ computed from 1000 samples of size $n = 128$ from models D(2), D(3), S(0.3), S(0.9), DS(2,0.4) and DS(4,0.7)

n increases, as expected theoretically. To convey some intuitions on the influence on h on the empirical power under non-stationarity, we again consider the same setup as before and plot the rejection percentages of c and dc computed from 1000 samples of size $n = 128$ from models D(2), D(3), S(0.3), S(0.9), DS(2,0.4) and DS(4,0.7) against the embedding dimension h. As one can see from Figure 1, for the models under consideration, the empirical powers of the tests c and dc essentially decrease as h increases. Additional simulations presented in Appendix S1 and involving an AR(2) model instead of an AR(1) model for the serial dependence show that a similar pattern occurs from $h = 3$ onward in that case. Indeed, as discussed in Section 3.4, for many models including those that were just mentioned, the power of the tests appears to be a decreasing function of h, at least from some small value of h onward.

How do the studied tests compare to existing competitors? As mentioned in the introduction, many tests of stationarity were proposed in the literature. Unfortunately, only a few of them seem to be implemented in statistical software. In Appendix S1, we report the results of Monte Carlo experiments investigating the finite-sample behavior of the tests of Priestley and Subba Rao (1969), Nason (2013) and Cardinali and Nason (2013) that are implemented in the R packages `fractal` (Constantine and Percival, 2016), `locits` (Nason, 2016) and `costat` (Nason and Cardinali, 2013) respectively. Note that we did not consider the test of Cardinali and Nason (2016) (implemented in the R package `BootWPTOS`) because we were unable to understand how to initialize the arguments of the corresponding R function. Under stationarity, unlike the rank-based tests d, c, dc and dcp, the three aforementioned tests were found to be too liberal for at least one of the considered models. Their behavior under the null turned out to be even more disappointing when heavy-tailed innovations were used. In terms of empirical

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

Table III. Approximate p-values (multiplied by 100) of the rank-based tests of stationarity proposed in Section 3 for embedding dimension $h \in \{2, 3, 4\}$ applied to the component times series of the trivariate log-return data considered in McNeil *et al.* (2005, Chapter 5) and the bivariate log-return data considered in Grégoire *et al.* (2008). The daily log-returns of the Intel, Microsoft and General Electric stocks are abbreviated by INTC, MSFT and GE respectively. The meanings of the abbreviations d, c, dc and dcp are given in Section 5. The columns c2 and c3 report the results for the bivariate analogues of the test based on $S_{n,C^{(2)}}$ defined by (3.3) (which arise in the combined test dcp) for lags 2 and 3

Model		$h = 2$ or lag 1		$h = 3$ or lag 2				$h = 4$ or lag 3			
Variable	d	c	dc	c	dc	c2	dcp	c	dc	c3	dcp
INTC	0.0	2.0	0.0	4.8	0.0	32.5	0.0	7.9	0.0	30.2	0.0
MSFT	0.2	92.3	2.2	80.7	0.8	47.3	0.0	86.4	0.1	37.2	0.0
GE	0.1	62.1	0.7	15.9	0.1	67.2	0.0	22.4	0.6	16.7	0.1
Oil	89.6	22.1	52.5	55.3	84.0	46.5	67.8	89.0	97.2	5.6	49.0
Gas	5.0	16.5	3.9	17.4	5.4	90.5	7.4	43.9	8.8	85.2	6.2

power, the results presented in Appendix S1 allow, in principle, for a direct comparison with the results reported in Cardinali and Nason (2013) and Dette *et al.* (2011). Since the tests available in R considered in Cardinali and Nason (2013) are far from maintaining their levels, a comparison in terms of power with these tests is clearly not meaningful. As far as the tests of Dette *et al.* (2011) are concerned, they appear, overall, to be more powerful for some of the considered models. It is, however, unknown whether they hold their levels when applied to stationary heavy-tailed observations as only Gaussian time series were considered in the simulations of Dette *et al.* (2011).

6. ILLUSTRATIONS

By construction, the tests based on the sample mean, variance and autocovariance proposed in Section 4 are only sensitive to changes in the second-order characteristics of a time series. The results of the simulations reported in the previous section and in Appendix S1 seem to indicate that the latter tests do not always maintain their level (for instance, in the presence of conditional heteroskedasticity) and that the rank-based tests proposed in Section 3 are more powerful, even in situations only involving changes in the second-order characteristics. Therefore, we recommend the use of the rank-based tests in general.

To illustrate their application, we consider two real datasets, both available in the R package `copula` (Hofert *et al.*, 2017). The first one consists of daily log-returns of Intel, Microsoft and General Electric stocks for the period from 1996 to 2000. It was used in McNeil *et al.* (2005, Chapter 5) to illustrate the fitting of elliptical copulas. The second dataset was initially considered in Grégoire *et al.* (2008) to illustrate the so-called *copula–GARCH* approach (see, e.g., Chen and Fan, 2006; Patton, 2006). It consists of bivariate daily log-returns computed from 3 years of daily prices of crude oil and natural gas for the period from July 2003 to July 2006.

Prior to applying the methodologies described in the aforementioned references, it is crucial to assess whether the available data can be regarded as stretches from stationary multi-variate time series. As multi-variate versions of the proposed tests would need to be thoroughly investigated first (see the discussion in the next section), as an imperfect alternative, we applied the studied univariate versions to each component time series. The results are reported in Table III. For the sake of simplicity, we shall ignore the necessary adjustment of p-values or global significance level due to multiple testing.

As one can see from the results of the combined tests dc and dcp for embedding dimension $h \in \{2, 3, 4\}$, there is strong evidence against stationarity in the component series of the trivariate log-return data considered in McNeil *et al.* (2005, Chapter 5). For all three series, the very small p-values of the combined tests are a consequence of the very small p-value of the test d focusing on the contemporary distribution. For the Intel stock (line INTC), it is also a consequence of the small p-value of the test c for $h = 2$. Although it is, for instance, very tempting to conclude that the non-stationarity in the log-returns of the Intel stock is due to $H_0^{(1)}$ in (1.3) and $H_{0,c}^{(2)}$ in (1.4) not being satisfied, such a reasoning is not formally valid without additional assumptions, as explained in the introduction.

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

From the second horizontal block of Table III, one can also conclude that there is no evidence against stationarity in the log-returns of the oil prices and only weak evidence against stationarity in the log-returns of the gas prices.

7. CONCLUDING REMARKS

Unlike some of their competitors that are implemented in various R packages, the rank-based tests of stationarity proposed in Section 3 were never observed to be too liberal for the rather typical sample sizes considered in this work. As discussed in Section 3.4, and as empirically confirmed by the experiments of Section 5 and Appendix S1, the tests are nevertheless likely to become more conservative and less powerful as the embedding dimension h is increased. The latter led us to make the rather general recommendation that they should be typically used with a small value of the embedding dimension h such as 2, 3 or 4. It is, however, difficult to assess the breadth of that recommendation, and it might be meaningful for the practitioner to consider the issue of the choice of h in all its subtlety as attempted in the discussion of Section 3.4.

While, unsurprisingly, the recommended tests seem to display good power for alternatives connected to the change-point detection literature, their power was not observed to be very high, overall, for the locally stationary alternatives considered in our Monte Carlo experiments. A promising approach to improve on the latter aspect would be to derive extensions of the tests allowing the comparison of blocks of observations in the spirit of Hušková and Slabý (2001) and of Eichinger and Kirch (2018): once the time series is divided into moving blocks of equal length, the main idea is to compare successive pairs of blocks by means of a statistic based on a suitable extension of the process in (3.4) (if the focus is on serial dependence) or in (3.16) (if the focus is on the contemporary distribution) and to finally aggregate the statistics for each pair of blocks.

Additional future research may consist of extending the proposed tests to multi-variate time series. To fix ideas, let us focus on lag $h-1$ and consider a stretch $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,d})$, $i \in \{1, \dots, n+h-1\}$ from a continuous d-dimensional time series. A straightforward extension of the approach considered in this work is first to define the $d \times h$-dimensional random vectors $\boldsymbol{Y}_i^{(h)} = (\boldsymbol{X}_i, \dots, \boldsymbol{X}_{i+h-1})$, $i \in \{1, \dots, n\}$. As argued in the introduction and in Section 5, it will then be helpful in terms of finite-sample power properties to split the hypothesis $H_0^{(h)}$ in (1.2) into suitable subhypotheses. For $A \subset \{1, \dots, d\}$ and $B \subset \{0, \dots, h-1\}$, let

$$
\begin{aligned}
H_0^{(1)}(A) &: \exists\, G^A \text{ such that } (X_{1,j})_{j\in A}, \dots, (X_{n-h+1,j})_{j\in A} \text{ have d.f. } G^A,\\
H_{0,c}^{(h)}(A,B) &: \exists\, C^{(h),A,B} \text{ such that } (X_{1+s,j})_{s\in B, j\in A}, \dots, (X_{n+s,j})_{s\in B, j\in A} \text{ have copula } C^{(h),A,B}.
\end{aligned}
$$

Letting $\bar{d} = \{1, \dots, d\}$ and $\bar{h} = \{0, \dots, h-1\}$, Sklar's theorem suggests the decomposition of $H_0^{(h)} = H_0^{(1)}(\{1\}) \cap \cdots \cap H_0^{(1)}(\{d\}) \cap H_{0,c}^{(h)}(\bar{d}, \bar{h})$. However, preliminary numerical experiments indicate that a straightforward extension of the approach proposed in Section 3.3 to this combined hypothesis does not seem to be very powerful. The latter might be due to the curse of dimensionality identified in Section 3.4 and the fact that, under stationarity, the $d \times h$-dimensional copula $C^{(h),\bar{d},\bar{h}}$ of the $\boldsymbol{Y}_i^{(h)}$ arising in the aforementioned decomposition does not solely control the serial dependence in the time series but also the cross-sectional dependence. As a consequence, alternative combination strategies would need to be investigated in the multi-variate case. As an imperfect alternative, one might, for instance, consider the following hypothesis

$$
\Big(\bigcap_{j=1}^d H_0^{(1)}(\{j\}) \Big) \cap \Big(H_{0,c}^{(h)}(\bar{d}, \{0\}) \Big) \cap \Big(\bigcap_{j=1}^d H_{0,c}^{(h)}(\{j\}, \bar{h}) \Big),
$$

a combined test of which would be sensible to any changes in the marginals, the contemporary dependence or the marginal serial dependence. One may easily include further hypotheses related to cross-sectional and cross-serial dependencies, such as $\bigcap_{i \neq j \in \bar{d}} H_{0,c}^{(h)}(\{i,j\}, \{0,1\})$. The amount of potential adaptations appears to be very large, whence a further investigation, particularly from a finite-sample point of view, is beyond the scope of this paper.

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

ACKNOWLEDGMENTS

The authors thank two anonymous referees and a coeditor for their constructive and insightful comments on an earlier version of this manuscript. Axel Bücher gratefully acknowledges support by the Collaborative Research Center 'Statistical modeling of nonlinear dynamic processes' (SFB 823) of the German Research Foundation. Parts of this article were written when Axel Bücher was a postdoctoral researcher at Ruhr-Universität Bochum, Germany. Jean-David Fermanian's work was supported by the grant 'Investissements d'Avenir' (ANR-11-IDEX0003/Labex Ecodec) of the French National Research Agency.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Aue A, Horváth L. 2013. Structural breaks in time series. *Journal of Time Series Analysis* **34**(1): 1–16.

Bücher A. 2015. A note on weak convergence of the sequential multi-variate empirical process under strong mixing. *Journal of Theoretical Probability* **28**(3): 1028–1037.

Bücher A, Kojadinovic I. 2016a. Dependent multiplier bootstraps for non-degenerate *u*-statistics under mixing conditions with applications. *Journal of Statistical Planning and Inference* **170**: 83–105.

Bücher A, Kojadinovic I. 2016b. A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli* **22**(2).

Bücher A, Kojadinovic I. 2018. A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*. In press.

Bücher A, Kojadinovic I, Rohmer T, Segers J. 2014. Detecting changes in cross-sectional dependence in multi-variate time series. *Journal of Multi-variate Analysis* **132**: 111–128.

Bücher A, Volgushev S. 2013. Empirical and sequential empirical copula processes under serial dependence. *Journal of Multi-variate Analysis* **119**: 61–70.

Cardinali A, Nason G. 2013. Costationarity of locally stationary time series using costat. *Journal of Statistical Software* **55**(1): 1–22.

Cardinali A, Nason G. 2016. Practical powerful wavelet packet tests for second-order stationarity. *Applied and Computational Harmonic Analysis*. In press.

Chen X, Fan Y. 2006. Estimation and model selection of semiparametric copula-based multi-variate dynamic models under copula misspecification. *Journal of Econometrics* **135**: 125–154.

Chernozhukov V, Chetverikov D, Kato K. 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* **41**(6): 2786–2819.

Constantine W, Percival D. 2016. *fractal: Fractal Time Series Modeling and Analysis*. R package version 2.0-1.

Csörgö M, Horváth L. 1997. *Limit theorems in change-point analysis. Wiley Series in Probability and Statistics*. Chichester, UK: John Wiley and Sons.

Davydov YA, Lifshits MA. 1984. The fibering method in some probability problems. In *Probability Theory. Mathematical Statistics. Theoretical Cybernetics. Itogi Nauki i Tekhniki*, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform: Moscow, 61–157.

Dedecker J, Merlevède F, Rio E. 2014. Strong approximation of the empirical distribution function for absolutely regular sequences in. R_d. *Electronic Journal of Probability* **19**(9): 56.

Deheuvels P. 1979. La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance. *Bulletin de la Classe des Sciences. 5e Série* **65**: 274–292.

Deheuvels P. 1981. A non parametric test for independence. *Publications de l'Institut de Statistique de l'Université de Paris* **26**: 29–50.

Dereich S, Fehringer F, Matoussi A, Scheutzow M. 2003. On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *Journal of Theoretical Probability* **16**(1): 249–265.

Dette H, Preuss P, Vetter M. 2011. A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association* **106**: 1113–1124.

Dette H, Wu W, Zhou Z. 2015. Change point analysis of second order characteristics in non-stationary time series. arXiv:1801.10478v1.

Dwivedi Y, Subba Rao S. 2011. A test for second-order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis* **32**: 68–91.

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

Eichinger B, Kirch C. 2018. A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24**(1): 526–564.
Escanciano JC, Lobato IN. 2009. An automatic portmanteau test for serial correlation. *Journal of Econometrics* **151**(2): 140–149. Recent Advances in Time Series Analysis: A Volume Honouring Peter M. Robinson.
Fisher R. 1932. *Statistical Methods for Research Workers*. London: Olivier and Boyd.
Genest C, Rémillard B. 2004. Tests of independence and randomness based on the empirical copula process. *Test* **13**(2): 335–369.
Gombay E, Horváth L. 1999. Change-points and bootstrap. *Environmetrics* **10**(6).
Grégoire V, Genest C, Gendron M. 2008. Using copulas to model price dependence in energy markets. *Energy Risk* **5**(5): 58–64.
Hofert M, Kojadinovic I, Mächler M, Yan J. 2017. *copula: Multi-variate Dependence with Copulas*. R package version 0.999-17.
Holmes M, Kojadinovic I, Quessy J-F. 2013. Nonparametric tests for change-point detection à la Gombay and Horváth. *Journal of Multi-variate Analysis* **115**: 16–32.
Hušková M, Meintanis S. 2006a. Change-point analysis based on empirical characteristic functions. *Metrika* **63**: 145–168.
Hušková M, Meintanis S. 2006b. Change-point analysis based on empirical characteristic functions of ranks. *Sequential Analysis* **25**(4): 421–436.
Hušková M, Slabý A. 2001. Permutation tests for multiple changes. *Kybernetika (Prague)* **37**(5): 605–622.
Inoue A. 2001. Testing for distributional change in time series. *Econometric Theory* **17**(1): 156–187.
Jin L, Wang S, Wang H. 2015. A new non-parametric stationarity test of time series in the time domain. *Journal of the Royal Statistical Society: Series B* **77**: 893–922.
Kojadinovic I. 2017. *npcp: Some Nonparametric Tests for Change-Point Detection in Possibly Multi-variate Observations*. R package version 0.1-9.
Kosorok M. 2008. *Introduction to Empirical Processes and Semiparametric Inference*. New York, NY: Springer.
Künsch H. 1989. The jacknife and the bootstrap for general stationary observations. *The Annals of Statistics* **17**(3): 1217–1241.
Lee S, Ha J, Na O. 2003. The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics* **30**: 781–796.
McNeil A, Frey R, Embrechts P. 2005. *Quantitative Risk Management*. Princeton, NJ: Princeton University Press.
Nason G. 2013. A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society: Series B* **75**: 879–904.
Nason G. 2016. *locits: Tests of Stationarity and Localized Autocovariance*. R package version 1.7-3.
Nason G, Cardinali A. 2013. *costat: Time Series Costationarity Determination*. R package version 2.3.
Nelsen R. 2006. *An Introduction to Copulas*, 2nd ed. New York, NY: Springer.
Nualart D. 2006. Probability and its applications. In *The Malliavin Calculus and Related Topics*, Springer-Verlag: New York, Berlin.
Page E. 1954. Continuous inspection schemes. *Biometrika* **41**(1/2): 100–115.
Paparoditis E. 2010. Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association* **105**: 839–851.
Paparoditis E, Politis D. 2001. Tapered block bootstrap. *Biometrika* **88**(4): 1105–1119.
Patton A. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* **47**(2): 527–556.
Patton A, Politis D, White H. 2009. Correction: Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* **28**(4): 372–375.
Phillips P. 1987. Time series regression with unit roots. *Econometrica* **55**: 277–301.
Politis D, White H. 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* **23**(1): 53–70.
Preuss P, Vetter M, Dette H. 2013. A test for stationarity based on empirical processes. *Bernoulli* **19**(5B): 2715–2749.
Priestley M, Subba Rao T. 1969. A test for non-stationarity of time-series. *Journal of the Royal Statistical Society: Series B* **31**: 140–149.
Puchstein R, Preuss P. 2016. Testing for stationarity in multi-variate locally stationary processes. *Journal of Time Series Analysis* **37**(1): 3–29.
R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rüschendorf L. 1985. Construction of multi-variate distributions with given marginals. *Annals of the Institute of Statistical Mathematics* **37**(2): 225–233.
Segers J. 2012. Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli* **18**: 764–782.
Sklar A. 1959. Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**: 229–231.
Stouffer S, Suchman E, DeVinney L, Star S, Williams RJ. 1949. *The American Soldier. Adjustment during Army Life*, Vol. 1. Princeton, NJ: Princeton University Press.

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

van der Vaart A. 1998. *Asymptotic Statistics*. Cambridge: Cambridge University Press.
van der Vaart A, Wellner J. 2000. *Weak Convergence and Empirical Processes*, 2nd ed. New York: Springer.
von Sachs R, Neumann M. 2000. A wavelet-based test for stationarity. *Journal of Time Series Analysis* **21**: 597–613.
Zhou Z. 2013. Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association* **108**: 726–740.

APPENDIX A: DEPENDENT MULTIPLIER SEQUENCES

A sequence of random variables $(\xi_{i,n})_{i\in\mathbb{Z}}$ is a *dependent multiplier sequence* if the three following conditions are fulfilled:

($\mathcal{M}1$) The sequence $(\xi_{i,n})_{i\in\mathbb{Z}}$ is independent of the available sample $X_1,\dots,X_{n+h-1}$ and strictly stationary with $\mathrm{E}(\xi_{0,n})=0$, $\mathrm{E}(\xi_{0,n}^2)=1$ and $\sup_{n\geq 1}\mathrm{E}(|\xi_{0,n}|^\nu)<\infty$ for all $\nu\geq 1$.
($\mathcal{M}2$) There exists a sequence $\ell_n\to\infty$ of strictly positive constants such that $\ell_n=o(n)$ and the sequence $(\xi_{i,n})_{i\in\mathbb{Z}}$ is ℓ_n-dependent, that is, $\xi_{i,n}$ is independent of $\xi_{i+p,n}$ for all $p>\ell_n$ and $i\in\mathbb{N}$.
($\mathcal{M}3$) There exists a function $\varphi:\mathbb{R}\to[0,1]$, symmetric around 0, continuous at 0, satisfying $\varphi(0)=1$ and $\varphi(x)=0$ for all $|x|>1$ such that $\mathrm{E}(\xi_{0,n}\xi_{p,n})=\varphi(p/\ell_n)$ for all $p\in\mathbb{Z}$.

Roughly speaking, such sequences extend to the serially dependent setting, the multiplier sequences that appear in the *multiplier central limit theorem* (see, e.g., Kosorok, 2008, Theorem 10.1 and Corollary 10.3). The latter result lies at the heart of the proof of the asymptotic validity of many types of bootstrap schemes for independent observations. In particular, and as it shall become clearer below, the bandwidth parameter ℓ_n plays a role somewhat similar to the block length in the block bootstrap of Künsch (1989).

Two ways of generating dependent multiplier sequences are discussed in Bücher and Kojadinovic (2016b, Section 5.2). Throughout this work, we use the so-called *moving average approach* based on an initial i.i.d. standard normal sequence and Parzen's kernel

$$\kappa(x)=(1-6x^2+6|x|^3)\mathbf{1}(|x|\leq 1/2)+2(1-|x|)^3\mathbf{1}(1/2<|x|\leq 1),\quad x\in\mathbb{R}.$$

Specifically, let (b_n) be a sequence of integers such that $b_n\to\infty$, $b_n=o(n)$ and $b_n\geq 1$ for all $n\in\mathbb{N}$. Let $Z_1,\dots,Z_{n+2b_n-2}$ be i.i.d. $\mathcal{N}(0,1)$. Then, let $\ell_n=2b_n-1$, and for any $j\in\{1,\dots,\ell_n\}$, let $w_{j,n}=\kappa\{(j-b_n)/b_n\}$ and $\tilde{w}_{j,n}=w_{j,n}(\sum_{j'=1}^{\ell_n}w_{j',n}^2)^{-1/2}$. Finally, for all $i\in\{1,\dots,n\}$, let

$$\xi_{i,n}=\sum_{j=1}^{\ell_n}\tilde{w}_{j,n}Z_{j+i-1}. \tag{A.1}$$

Then, as verified in Bücher and Kojadinovic (2016b, Section 5.2), the infinite size version of $\xi_{1,n},\dots,\xi_{n,n}$ satisfies Assumptions ($\mathcal{M}1$)–($\mathcal{M}3$), when n is sufficiently large.

As can be expected, the bandwidth parameter ℓ_n (or, equivalently, b_n) will have a crucial influence on the finite-sample performance of the tests studied in this work. In practice, for the rank-based (resp. second-order) tests of Section 3 (resp. Section 4), we apply to the available univariate sequence $X_1,\dots,X_{n+h-1}$, the data-adaptive procedure proposed in Bücher and Kojadinovic (2016b, Section 5.1) (resp. Bücher and Kojadinovic, 2016a, Section 2.4), which is based on the seminal work of Paparoditis and Politis (2001), Politis and White (2004) and Patton *et al.* (2009), among others. Roughly speaking, the latter amounts to choosing ℓ_n as $K_n n^{1/5}$, which asymptotically minimizes a certain integrated mean squared error for a constant K_n that can be estimated from $X_1,\dots,X_{n+h-1}$.

Monte Carlo experiments studying the finite-sample behavior of the data-adaptive procedure of Bücher and Kojadinovic (2016b, Section 5.1) for estimating the bandwidth parameter b_n can be found in Bücher and Kojadinovic (2016b, Section 6). A small simulation showing how the automatically chosen bandwidth

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

parameter b_n is affected by the strength of the serial dependence in an AR(1) model is presented in Appendix S1.

APPENDIX B: PROOFS

Proof of Proposition 2.1. As we continue, we adopt the notation $\bar{F}^*_{T_j}(x) = \mathbb{P}(T^{[1]}_{n,j} \geq x \mid \boldsymbol{X}_n)$, $x \in \mathbb{R}$, $j \in \{1, \dots, r\}$. Note, in passing, that the functions $\bar{F}^*_{T_j}$ are random and that we can rewrite W_n in (2.6) as $W_n = \psi\{\bar{F}^*_{T_1}(T_{n,1}), \dots, \bar{F}^*_{T_r}(T_{n,r})\}$. In addition, recall that $\bar{F}_{T_j}(x) = \mathbb{P}(T_j \geq x)$, $x \in \mathbb{R}$, $j \in \{1, \dots, r\}$. Combining either (2.7) or (2.8) with Lemma 2.2 in Bücher and Kojadinovic (2018) and Problem 23.1 in van der Vaart (1998), we obtain:

$$\sup_{x\in\mathbb{R}} |\bar{F}^*_{T_j}(x) - \bar{F}_{T_j}(x)| \overset{\mathbb{P}}{\to} 0, \quad j \in \{1, \dots, r\}. \tag{B.1}$$

Furthermore, Lemma 2.2 in Bücher and Kojadinovic (2018) implies that (B.1) is equivalent to

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}(T^{[i]}_{n,j} \geq x) - \bar{F}_{T_j}(x) \right| \overset{\mathbb{P}}{\to} 0, \quad j \in \{1, \dots, r\}. \tag{B.2}$$

Again, from Lemma 2.2 in Bücher and Kojadinovic (2018), we also have that (2.7) or (2.8) imply that

$$(\boldsymbol{T}_n, \boldsymbol{T}^{[1]}_n, \dots, \boldsymbol{T}^{[N]}_n) \rightsquigarrow (\boldsymbol{T}, \boldsymbol{T}^{[1]}, \dots, \boldsymbol{T}^{[N]}),$$

for all $N \in \mathbb{N}$, where $\boldsymbol{T}^{[1]}, \dots, \boldsymbol{T}^{[N]}$ are independent copies of $\boldsymbol{T}$. Combining this last result with the continuous mapping theorem, we immediately obtain, for any $N \in \mathbb{N}$,

$$\left(\bar{\boldsymbol{F}}_T(\boldsymbol{T}_n), \bar{\boldsymbol{F}}_T\left(\boldsymbol{T}^{[1]}_n\right), \dots, \bar{\boldsymbol{F}}_T\left(\boldsymbol{T}^{[N]}_n\right)\right) \rightsquigarrow \left(\bar{\boldsymbol{F}}_T\left(\boldsymbol{T}\right), \bar{\boldsymbol{F}}_T\left(\boldsymbol{T}^{[1]}\right), \dots, \bar{\boldsymbol{F}}_T\left(\boldsymbol{T}^{[N]}\right)\right), \tag{B.3}$$

where $\bar{\boldsymbol{F}}_T(\boldsymbol{x}) = (\bar{F}_{T_1}(x_1), \dots, \bar{F}_{T_r}(x_r))$, $\boldsymbol{x} \in \mathbb{R}^r$. Combining (B.3) with (B.2), the continuity of ψ and the continuous mapping theorem, we observe that (2.9) holds for all $N \in \mathbb{N}$.

From now on, assume that W has a continuous d.f. As a straightforward consequence of (B.1) and the continuous mapping theorem, the weak convergence in (B.3) implies that, for any $N \in \mathbb{N}$,

$$\left(W_n, W^{[1]}_n, \dots, W^{[N]}_n\right) \rightsquigarrow \left(W, W^{[1]}, \dots, W^{[N]}\right),$$

where W_n is defined by (2.6) and $W^{[i]}_n = \psi\{\bar{F}^*_{T_1}(T^{[i]}_{n,1}), \dots, \bar{F}^*_{T_r}(T^{[i]}_{n,r})\}$, $i \in \{1, \dots, N\}$. The previous display has the following two consequences: first, by Problem 23.1 in van der Vaart (1998),

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(W_n \leq x) - \mathbb{P}(W \leq x)| \overset{\mathbb{P}}{\to} 0. \tag{B.4}$$

Second, since $W^{[1]}_n, \dots, W^{[N]}_n$ are identically distributed and conditionally independent on the data, using Lemma 2.2 in Bücher and Kojadinovic (2018), we have

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(W^{[1]}_n \leq x \mid \boldsymbol{X}_n\right) - \mathbb{P}\left(W_n \leq x\right)\right| \overset{\mathbb{P}}{\to} 0. \tag{B.5}$$

Next, let us prove (2.11). In view of (B.5), it suffices to show that

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(W^{[1]}_{n,M_n} \leq x \mid \boldsymbol{X}_n\right) - \mathbb{P}\left(W^{[1]}_n \leq x \mid \boldsymbol{X}_n\right)\right| \overset{\mathbb{P}}{\to} 0. \tag{B.6}$$

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

Using the fact that, for any $a, b, x \in \mathbb{R}$ and $\varepsilon > 0$,

$$|\mathbf{1}(a \le x) - \mathbf{1}(b \le x)| \le \mathbf{1}(|x-a| \le \varepsilon) + \mathbf{1}(|a-b| > \varepsilon), \tag{B.7}$$

we have

$$\begin{aligned}\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(W_n^{[1]} \le x \mid \boldsymbol{X}_n\right) - \mathbb{P}\left(W_{n,M_n}^{[1]} \le x \mid \boldsymbol{X}_n\right)\right| &\le \sup_{x\in\mathbb{R}} \mathbb{P}\left(\left|W_n^{[1]} - x\right| \le \varepsilon | \boldsymbol{X}_n\right) \\ &+ \mathbb{P}\left(\left|W_n^{[1]} - W_{n,M_n}^{[1]}\right| > \varepsilon \mid \boldsymbol{X}_n\right).\end{aligned}$$

From (B.4) and (B.5), $\sup_{x\in\mathbb{R}} \mathbb{P}(|W_n^{[1]} - x| \le \varepsilon \mid \boldsymbol{X}_n)$ converges in probability to $\sup_{x\in\mathbb{R}} \mathbb{P}(|W - x| \le \varepsilon)$, which can be made arbitrary small by decreasing ε. From (B.1)–(B.3) and the continuous mapping theorem, we observe that $W_n^{[1]} - W_{n,M_n}^{[1]} = o_{\mathbb{P}}(1)$, which implies that

$$\mathbb{P}\left(\left|W_n^{[1]} - W_{n,M_n}^{[1]}\right| > \varepsilon\right) = \mathrm{E}\left\{\mathbb{P}\left(\left|W_n^{[1]} - W_{n,M_n}^{[1]}\right| > \varepsilon \mid \boldsymbol{X}_n\right)\right\} \to 0, \tag{B.8}$$

and thus that $\mathbb{P}(|W_n^{[1]} - W_{n,M_n}^{[1]}| > \varepsilon \mid \boldsymbol{X}_n) = o_{\mathbb{P}}(1)$. Hence, (B.6) holds and thus so does (2.11).

Finally, let us show that (2.12) holds. Since $W_n^{[1]}, \dots, W_n^{[N]}$ are identically distributed and independent conditionally on the data, using Lemma 2.2 in Bücher and Kojadinovic (2018), we have that (B.5) implies

$$\sup_{x\in\mathbb{R}} \left|\frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(W_n^{[i]} \le x\right) - \mathbb{P}\left(W_n \le x\right)\right| \xrightarrow{\mathbb{P}} 0. \tag{B.9}$$

Whence (2.12) is proven if we show that

$$\sup_{x\in\mathbb{R}} \left|\frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(W_{n,M_n}^{[i]} \le x\right) - \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(W_n^{[i]} \le x\right)\right| \xrightarrow{\mathbb{P}} 0. \tag{B.10}$$

Again using (B.7), the term on the left of the previous display is smaller than

$$\sup_{x\in\mathbb{R}} \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(\left|W_n^{[i]} - x\right| \le \varepsilon\right) + \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(\left|W_n^{[i]} - W_{n,M_n}^{[i]}\right| \ge \varepsilon\right).$$

From (B.9) and (B.4), the first term converges in probability to $\sup_{x\in\mathbb{R}} \mathbb{P}(|W-x| \le \varepsilon)$, which can be made arbitrary small by decreasing ε. The second term converges in probability to zero by Markov's inequality: for any $\lambda > 0$,

$$\begin{aligned}\mathbb{P}\left\{\frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(\left|W_n^{[i]} - W_{n,M_n}^{[i]}\right| \ge \varepsilon\right) > \lambda\right\} &\le \lambda^{-1} \mathrm{E}\left\{\frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}\left(\left|W_n^{[i]} - W_{n,M_n}^{[i]}\right| \ge \varepsilon\right)\right\} \\ &\le \lambda^{-1} \mathbb{P}\left(\left|W_n^{[1]} - W_{n,M_n}^{[1]}\right| \ge \varepsilon\right) \to 0\end{aligned}$$

since the $W_n^{[i]} - W_{n,M_n}^{[i]}$ are identically distributed and by (B.8). Therefore, (B.10) holds and, hence, so does (2.12). Note that, from the fact that $\boldsymbol{T}$ and W have continuous d.f.s, we could have alternatively proven the analogue statement with '$\le$' replaced by '$<$'. As a consequence, we immediately find that $p_{n,M_n}(W_{n,M_n}^{[0]})$ has the same weak limit as $\bar{F}_{W_n}(W_{n,M_n}^{[0]})$, where $\bar{F}_{W_n}(w) = \mathbb{P}(W_n \ge w)$, $w \in \mathbb{R}$. Through the analogue to (B.4), with '$\le$' replaced by '$<$', the latter has the same asymptotic distribution as $\bar{F}_W(W_{n,M_n}^{[0]})$, where $\bar{F}_W(w) = \mathbb{P}(W \ge w)$, $w \in \mathbb{R}$. By the weak

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431

convergence $W^{[0]}_{n,M_n} \rightsquigarrow W$ following from (2.9) and the continuous mapping theorem, $\bar{F}_W(W^{[0]}_{n,M_n})$ is asymptotically standard uniform. ■

Proof of Proposition 2.2. Notice first that assumption (ii) implies that the corresponding approximate p-value $p_{n,M_n}(T^{[0]}_{n,j_0})$ given by (2.3) converges to zero in probability. Indeed,

$$\mathbb{E}\left\{p_{n,M_n}\left(T^{[0]}_{n,j_0}\right)\right\} = \frac{1}{M_n+1}\left\{\frac{1}{2} + \sum_{k=1}^{M_n} \mathbb{P}\left(T^{[k]}_{n,j_0} \geq T^{[0]}_{n,j_0}\right)\right\} = \mathbb{P}\left(T^{[1]}_{n,j_0} \geq T^{[0]}_{n,j_0}\right) + O\left(M_n^{-1}\right).$$

Next, a consequence of assumption (*iii*) is that, for any $j \in \{1, \dots, r\}$,

$$\left(p_{n,M_n}\left(T^{[1]}_{n,j}\right), \dots, p_{n,M_n}\left(T^{[M_n]}_{n,j}\right)\right)$$

is a permutation of the vector

$$\left(\tfrac{3}{2M_n+2}, \dots, \tfrac{2M_n+1}{2M_n+2}\right).$$

It follows that, for any $x \in (0,1)$,

$$\frac{1}{M_n}\sum_{k=1}^{M_n} \mathbf{1}\left\{p_{n,M_n}\left(T^{[k]}_{n,j}\right) \leq x\right\} = \frac{1}{M_n}\sum_{k=1}^{M_n} \mathbf{1}\left(\tfrac{2k+1}{2M_n+2} \leq x\right) = x + O\left(M_n^{-1}\right), \tag{A.11}$$

where $\lfloor . \rfloor$ is the floor function. Then, let $\bar{w} = \max_{j\in\{1,\dots,r\}} w_j$. Starting from (2.5), and relying on assumptions (*i*) and (*iii*), we successively obtain

$$\begin{aligned}
p_{n,M_n}(W^{[0]}_{n,M_n}) &= \frac{1}{M_n}\sum_{k=1}^{M_n}\mathbf{1}\Big[\sum_{j=1}^{r} w_j\varphi\{p_{n,M_n}(T^{[k]}_{n,j})\} \geq \sum_{j=1}^{r} w_j\varphi\{p_{n,M_n}(T^{[0]}_{n,j})\}\Big] \\
&\leq \frac{1}{M_n}\sum_{k=1}^{M_n}\mathbf{1}\Big[\bar{w}\sum_{j=1}^{r} \varphi\{p_{n,M_n}(T^{[k]}_{n,j})\} \geq w_{j_0}\varphi\{p_{n,M_n}(T^{[0]}_{n,j_0})\}\Big] \\
&\leq \frac{1}{M_n}\sum_{k=1}^{M_n}\mathbf{1}\Big[\exists j \in \{1,\dots,r\} : r\bar{w}\varphi\{p_{n,M_n}(T^{[k]}_{n,j})\} \geq w_{j_0}\varphi\{p_{n,M_n}(T^{[0]}_{n,j_0})\}\Big] \\
&\leq \frac{1}{M_n}\sum_{k=1}^{M_n}\sum_{j=1}^{r}\mathbf{1}\Big[\varphi\{p_{n,M_n}(T^{[k]}_{n,j})\} \geq \tfrac{w_{j_0}}{r\bar{w}}\varphi\{p_{n,M_n}(T^{[0]}_{n,j_0})\}\Big] \\
&= \frac{r}{M_n}\sum_{k=1}^{M_n}\mathbf{1}\Big[\varphi\big(\tfrac{2k+1}{2M_n+2}\big) \geq \tfrac{w_{j_0}}{r\bar{w}}\varphi\{p_{n,M_n}(T^{[0]}_{n,j_0})\}\Big] \\
&= \frac{r}{M_n}\sum_{k=1}^{M_n}\mathbf{1}\left(\tfrac{2k+1}{2M_n+2} \leq \varphi^{-1}\Big[\tfrac{w_{j_0}}{r\bar{w}}\varphi\{p_{n,M_n}(T^{[0]}_{n,j_0})\}\Big]\right) \\
&= r\varphi^{-1}\Big[\tfrac{w_{j_0}}{r\bar{w}}\varphi\{p_{n,M_n}(T^{[0]}_{n,j_0})\}\Big] + O\left(M_n^{-1}\right) \xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

where the last statement follows from (A.11) and the fact that $p_{n,M_n}(T^{[0]}_{n,j_0}) \xrightarrow{\mathbb{P}} 0$. ■

 © 2018 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12431

Proof of Proposition 3.2. The result is a consequence of Proposition 3.3 in Bücher *et al.* (2014) and the fact that the strong mixing coefficients of the sequence $(\boldsymbol{Y}_i^{(h)})_{i\in\mathbb{Z}}$ defined through (1.1) can be expressed from those of the sequence $(X_i)_{i\in\mathbb{Z}}$ as $\alpha_r^{\boldsymbol{Y}} = \alpha_{(r-h+1)\vee 0}^{X}$, $r \in \mathbb{N}$, where $\vee$ is the maximum operator. ■

Proof of Proposition 3.3. The assertions concerning weak convergence are simple consequences of the continuous mapping theorem and Proposition 3.2. It remains to be shown that $\mathcal{L}(S_{C^{(h)}})$, the distribution of $S_{C^{(h)}}$, is absolutely continuous with respect to the Lebesgue measure. For that purpose, note that, with probability one, the sample paths of $\mathbb{D}_{C^{(h)}}$ are elements of $\mathcal{C}([0,1]\times[0,1]^h)$, the space of continuous real-valued functions on $[0,1]\times[0,1]^h$. We may write $S_{C^{(h)}} = \{f(\mathbb{D}_{C^{(h)}})\}^2$, where

$$f : \mathcal{C}([0,1]^{h+1}) \to \mathbb{R}, \quad f(g) = \sup_{s\in[0,1]} \left\{ \int_{[0,1]^h} g^2(s,\boldsymbol{u})\,\mathrm{d}C^{(h)}(\boldsymbol{u}) \right\}^{1/2},$$

and it is sufficient to show that $\mathcal{L}\{f(\mathbb{D}_{C^{(h)}})\}$ is absolutely continuous. Now, if $\mathcal{C}([0,1]^{h+1})$ is equipped with the supremum norm $\|\cdot\|_\infty$, then f is continuous and convex. We may thus apply Theorem 7.1 in Davydov and Lifshits (1984): $\mathcal{L}\{f(\mathbb{D}_{C^{(h)}})\}$ is concentrated on $[a_0,\infty)$ and absolutely continuous on (a_0,∞), where

$$a_0 = \inf\{f(g) : g \text{ belongs to the support of } \mathcal{L}(\mathbb{D}_{C^{(h)}})\}.$$

It remains to be shown that $\mathcal{L}\{f(\mathbb{D}_{C^{(h)}})\}$ has no atom at a_0. First, note that $a_0 = 0$. Indeed, by Lemma 1.2(e) in Dereich *et al.* (2003), we have $\mathbb{P}(\|\mathbb{D}_{C^{(h)}}\|_\infty \le \varepsilon) > 0$ for any $\varepsilon > 0$. Hence, for any $\varepsilon > 0$, there exist functions g in the support of the distribution of $\mathbb{D}_{C^{(h)}}$ such that $f(g) \le \varepsilon$, whence $a_0 = 0$ as asserted. Moreover, $f(\mathbb{D}_{C^{(h)}}) = 0$ holds if and only if $\mathbb{D}_{C^{(h)}}(s,\boldsymbol{u}) = 0$ for any $s \in [0,1]$ and any $\boldsymbol{u}$ in the support of the distribution induced by $C^{(h)}$ (by continuity of the sample paths). Then, choose an arbitrary point $\boldsymbol{u}^*$ in the latter support such that $\sigma^2 = \mathrm{Var}\{\mathbb{C}_{C^{(h)}}(0,1,\boldsymbol{u}^*)\} > 0$. A straightforward calculation shows that $\mathbb{C}_{C^{(h)}}(0,1/2,\boldsymbol{u}^*)$ and $\mathbb{C}_{C^{(h)}}(1/2,1,\boldsymbol{u}^*)$ are uncorrelated and have the same variance $\frac{1}{2}\sigma^2$. Hence,

$$\begin{aligned}\mathrm{Var}\left\{\mathbb{D}_{C^{(h)}}\left(\tfrac{1}{2},\boldsymbol{u}^*\right)\right\} &= \mathrm{Var}\left\{\tfrac{1}{2}\mathbb{C}_{C^{(h)}}(0,1/2,\boldsymbol{u}^*) - \tfrac{1}{2}\mathbb{C}_{C^{(h)}}(1/2,1,\boldsymbol{u}^*)\right\}\\ &= \tfrac{1}{4}\mathrm{Var}\left\{\mathbb{C}_{C^{(h)}}(0,1/2,\boldsymbol{u}^*)\right\} + \tfrac{1}{4}\mathrm{Var}\left\{\mathbb{C}_{C^{(h)}}(1/2,1,\boldsymbol{u}^*)\right\} = \tfrac{1}{4}\sigma^2 > 0.\end{aligned}$$

As a consequence, $\mathbb{P}(f(\mathbb{D}_{C^{(h)}}) = 0) \le \mathbb{P}(\mathbb{D}_{C^{(h)}}(\frac{1}{2},\boldsymbol{u}^*) = 0) = 0$, which finally implies that $\mathcal{L}(f(\mathbb{D}_{C^{(h)}}))$, and therefore, $\mathcal{L}(S^{(h)})$ is absolutely continuous. ■

Proof of Proposition 3.4. The result is a consequence of Proposition 4.2 in Bücher *et al.* (2014) and the fact that the strong mixing coefficients of the sequence $(\boldsymbol{Y}_i^{(h)})_{i\in\mathbb{Z}}$ can be expressed as $\alpha_r^{\boldsymbol{Y}} = \alpha_{(r-h+1)\vee 0}^{X}$, $r \in \mathbb{N}$. ■

Proof of Proposition 3.5. The assertions concerning weak convergence are simple consequences of Theorem 1 of Bücher (2015) and of the continuous mapping theorem. Absolute continuity of S_G can be shown along similar lines as for $S_{C^{(h)}}$ in Proposition 3.3. ■

Proof of Proposition 3.6. To prove the first claim, one first needs to show that the finite-dimensional distributions of $\left(\mathbb{D}_{n,C^{(h)}}, \hat{\mathbb{D}}_{n,C^{(h)}}^{[1]}, \dots, \hat{\mathbb{D}}_{n,C^{(h)}}^{[M]}, \mathbb{E}_n, \hat{\mathbb{E}}_n^{[1]}, \dots, \hat{\mathbb{E}}_n^{[M]}\right)$ converge weakly to those of $\left(\mathbb{D}_{C^{(h)}}, \mathbb{D}_{C^{(h)}}^{[1]}, \dots, \mathbb{D}_{C^{(h)}}^{[M]}, \mathbb{E}, \mathbb{E}^{[1]}, \dots, \mathbb{E}^{[M]}\right)$. The proof is a more notationally involved version of the proof of Lemma A.1 in Bücher and Kojadinovic (2016b). Joint asymptotic tightness follows from Proposition 3.4 as well as from the fact that, for any $m \in \mathbb{N}$, $\hat{\mathbb{E}}_n^{[m]} \rightsquigarrow \mathbb{E}^{[m]}$ in $\ell^\infty([0,1]\times\mathbb{R})$ as a consequence of Corollary 2.2 in Bücher and Kojadinovic (2016b) and the continuous mapping theorem. ■

© 2018 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12431