SFB 823

Combining cumulative sum change-point detection tests for assessing the stationarity of univariate time series

Axel Bücher, Jean-David Fermanian, Ivan Kojadinovic

Discussion Paper

Nr. 16/2017

Combining cumulative sum change-point detection tests for assessing the stationarity of univariate time series

Axel Bücher [*] Jean-David Fermanian [†] Ivan Kojadinovic [‡]

September 13, 2017

Abstract

We derive tests of stationarity for continuous univariate time series by combining change-point tests sensitive to changes in the contemporary distribution with tests sensitive to changes in the serial dependence. Rank-based cumulative sum tests based on the empirical distribution function and on the empirical autocopula at a given lag are considered first. The combination of their dependent p-values relies on a joint dependent multiplier bootstrap of the two underlying statistics. Conditions under which the proposed combined testing procedure is asymptotically valid under stationarity are provided. After discussing the choice of the maximum lag to investigate, extensions based on tests solely focusing on second-order characteristics are proposed. The finite-sample behaviors of all the derived statistical procedures are investigated in large-scale Monte Carlo experiments and illustrations on two real data sets are provided. Extensions to multivariate time series are briefly discussed as well.

Keywords: copula, dependent p-value combination, multiplier bootstrap, rank-based statistics, tests of stationarity.

MSC 2010: 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 Lee et al. (2003), Dwivedi and Subba Rao (2011), Jin et al. (2015) and Dette et al. (2015). In particular, the

[*] Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstr. 150, 44780 Bochum, Germany. E-mail: axel.buecher@rub.de

[†] CREST-ENSAE, J120, 3, avenue Pierre-Larousse, 92245 Malakoff cedex, France. E-mail: jean-david.fermanian@ensae.fr

[‡] CNRS / Université de Pau et des Pays de l'Adour, Laboratoire de mathématiques et applications – IPRA, UMR 5142, B.P. 1155, 64013 Pau Cedex, France. E-mail: ivan.kojadinovic@univ-pau.fr

first and the last reference also clearly pertain to the change-point detection literature (see, e.g., Csörgő 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 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 $2 \le h \le N$, 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) the CUSUM test based on differences of empirical d.f.s studied in Gombay and Horváth (1999), Inoue (2001) and Holmes et al. (2013) (see also Section 2.2 below). However, preliminary numerical experiments revealed the low power of such an adaptation, 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 (that is, 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

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

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.

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

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

and

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

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 employed only with respect to second-order characteristics of a time series: see, e.g., 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)) can be carried out using the same type of resampling, their p-values can be combined. Another contribution of this work is thus to propose a global test of stationarity based on appropriate extensions to dependent tests of well-known p-value combination methods such as those of Fisher (1932) or Stouffer et al. (1949). The larger the embedding dimension h, the more the resulting combined test could be regarded as a general test of stationarity.

An interesting and desirable feature of the resulting testing procedure is that it is rank-based. 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 combined 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 (2016b): 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.

The paper is organized as follows. The detailed description of the combined rank-based test involving empirical d.f.s and empirical autocopulas is given in Section 2, 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. 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, in particular, discussing multivariate extensions of the proposed tests.

Auxiliary results and all proofs are deferred to a sequence of appendices. The studied tests are implemented in the package `npcp` (Kojadinovic, 2017) for the R statistical system (R Core Team, 2017). In the rest of the paper, the arrow '$\leadsto$' denotes weak convergence in the sense of Definition 1.3.3 in van der Vaart and Wellner (2000), while the arrow '$\overset{\mathrm{P}}{\to}$' denotes convergence in probability. All convergences are for $n \to \infty$ if not mentioned otherwise. Finally, given a set T, $\ell^\infty(T)$ denotes the space of all bounded real-valued functions on T equipped with the uniform metric.

2 A rank-based combined test sensitive to departures from $H_0^{(h)}$

The aim of this section is to derive a 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 departures from $H_{0,c}^{(h)}$ in (1.4). We start with the latter.

2.1 A copula-based test senstive 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

$$C_{k:l}^{(h)}(\boldsymbol{u}) = \frac{1}{l-k+1} \sum_{i=k}^{l} \prod_{j=1}^{h} \mathbf{1}\{G_{k:l}(X_{i+j-1}) \leq u_j\}, \qquad \boldsymbol{u} \in [0,1]^h, \tag{2.1}$$

where

$$G_{k:l}(x) = \frac{1}{l+h-k} \sum_{j=k}^{l+h-1} \mathbf{1}(X_j \leq x), \qquad x \in \mathbb{R}, \tag{2.2}$$

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.

2.1.1 Test statistic

The CUSUM statistic that we consider is

$$S_n^{(h)} = \sup_{s \in [0,1]} \int_{[0,1]^h} \left\{\mathbb{D}_n^{(h)}(s, \boldsymbol{u})\right\}^2 \mathrm{d}C_{1:n}^{(h)}(\boldsymbol{u}) = \max_{1 \leq k \leq n-1} \int_{[0,1]^h} \left\{\mathbb{D}_n^{(h)}(k/n, \boldsymbol{u})\right\}^2 \mathrm{d}C_{1:n}^{(h)}(\boldsymbol{u}), \tag{2.3}$$

where $\lfloor . \rfloor$ is the floor function,

$$\mathbb{D}_n^{(h)}(s, \boldsymbol{u}) = \sqrt{n}\lambda_n(0, s)\lambda_n(s, 1)\left\{C_{1:\lfloor ns \rfloor}^{(h)}(\boldsymbol{u}) - C_{\lfloor ns \rfloor+1:n}^{(h)}(\boldsymbol{u})\right\}, \qquad (s, \boldsymbol{u}) \in [0, 1]^{h+1}, \tag{2.4}$$

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 $S_n^{(h)}$. At the opposite, large values of $S_n^{(h)}$ provide evidence of non-stationarity. The coefficient $\sqrt{n}\lambda_n(0, s)\lambda_n(s, 1)$ in (2.4) is the classical normalizing term of the CUSUM approach. It ensures that, under suitable conditions, $S_n^{(h)}$ converges in distribution under the null hypothesis of stationarity. Analogously to what was explained in the introduction, the test based on $S_n^{(h)}$ should not be used to reject $H_{0,c}^{(h)}$ in (1.4) as its null hypothesis is stationarity.

2.1.2 Limiting null distribution

The limiting null distribution of $S_n^{(h)}$ turns out to be a corollary of a recent result by Bücher and Kojadinovic (2016a) and Bücher et al. (2014). Under $H_0^{(h)}$ in (1.2), it can be verified that $\mathbb{D}_n^{(h)}$ in (2.4) can be written as

$$\mathbb{D}_n^{(h)}(s, \boldsymbol{u}) = \lambda_n(s, 1)\,\mathbb{C}_n^{(h)}(0, s, \boldsymbol{u}) - \lambda_n(0, s)\,\mathbb{C}_n^{(h)}(s, 1, \boldsymbol{u}), \qquad (s, \boldsymbol{u}) \in [0, 1]^{h+1}, \tag{2.5}$$

where

$$\mathbb{C}_n^{(h)}(s, t, \boldsymbol{u}) = \sqrt{n}\,\lambda_n(s, t)\,\{C_{\lfloor ns \rfloor+1:\lfloor nt \rfloor}^{(h)}(\boldsymbol{u}) - C^{(h)}(\boldsymbol{u})\}, \qquad (s, t, \boldsymbol{u}) \in \Delta \times [0, 1]^h. \tag{2.6}$$

Hence, the null weak limit of the empirical process $\mathbb{D}_n^{(h)}$ follows from that of $\mathbb{C}_n^{(h)}$, which is called 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 2.1. *For any $j \in \{1, \ldots, h\}$, the partial derivative $\dot{C}_j^{(h)} = \partial C^{(h)}/\partial u_j$ exists and is continuous on $V_j^{(h)} = \{\boldsymbol{u} \in [0, 1]^h : u_j \in (0, 1)\}$.*

Condition 2.1 is nonrestrictive in the sense that it is necessary so that the candidate weak limit of $\mathbb{C}_n^{(h)}$ exists pointwise and has continuous sample paths. In the sequel, following Bücher and Volgushev (2013), for any $j \in \{1, \ldots, h\}$, we define $\dot{C}_j^{(h)}$ to be zero on the set $\{\boldsymbol{u} \in [0, 1]^h : u_j \in \{0, 1\}\}$. Also, as we continue, for any $j \in \{1, \ldots, h\}$ and any $\boldsymbol{u} \in [0, 1]^h$, $\boldsymbol{u}^{(j)}$ will stand for the vector of $[0, 1]^h$ defined by $u_i^{(j)} = u_j$ if $i = j$ and 1 otherwise.

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

$$\mathbb{B}_n^{(h)}(s, t, \boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \left[\prod_{j=1}^{h} \mathbf{1}\{G(X_{i+j-1}) \leq u_j\} - C^{(h)}(\boldsymbol{u})\right], \qquad (s, t, \boldsymbol{u}) \in \Delta \times [0, 1]^h, \tag{2.7}$$

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

The following result, stating the weak limit of $\mathbb{C}_n^{(h)}$ and proved in Appendix B, is a consequence of the results of Bücher and Kojadinovic (2016a) and Bücher et al. (2014). It considers $X_1, \ldots, X_{n+h-1}$ 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 $(Z_i)_{a \leq i \leq b}$, $a, b \in \mathbb{Z} \cup \{-\infty, +\infty\}$, is denoted by $\mathcal{F}_a^b$. The strong mixing coefficients corresponding to the sequence $(Z_i)_{i \in \mathbb{Z}}$ are then defined by $\alpha_0^Z = 1/2$,

$$\alpha_r^Z = \sup_{p \in \mathbb{Z}} \sup_{A \in \mathcal{F}_{-\infty}^p, B \in \mathcal{F}_{p+r}^{+\infty}} \big|\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)\big|, \qquad r \in \mathbb{N}, r > 0. \tag{2.8}$$

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

Proposition 2.2. *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 $\alpha_r^X = O(r^{-a})$ for some $a > 1$ as $r \to \infty$. Then, provided Condition 2.1 holds,*

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

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

$$\mathbb{C}_C^{(h)}(s,t,\boldsymbol{u}) = \mathbb{B}_C^{(h)}(s,t,\boldsymbol{u}) - \sum_{j=1}^h \dot{C}_j^{(h)}(\boldsymbol{u})\, \mathbb{B}_C^{(h)}(s,t,\boldsymbol{u}^{(j)}), \tag{2.9}$$

and $\mathbb{B}_C^{(h)}$ in $\ell^\infty(\Delta\times[0,1]^h)$, a tight centered Gaussian process, is the weak limit of $\mathbb{B}_n^{(h)}$ in (2.7).

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 above mentioned references.

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

Proposition 2.3. *Under the conditions of Proposition 2.2, $\mathbb{D}_n^{(h)} \rightsquigarrow \mathbb{D}_C^{(h)}$ in $\ell^\infty([0,1]^{h+1})$, where, for any $(s,\boldsymbol{u}) \in [0,1]^{h+1}$,*

$$\mathbb{D}_C^{(h)}(s,\boldsymbol{u}) = \mathbb{C}_C^{(h)}(0,s,\boldsymbol{u}) - s\,\mathbb{C}_C^{(h)}(0,1,\boldsymbol{u}), \tag{2.10}$$

and $\mathbb{C}_C^{(h)}$ is defined in (2.9). As a consequence, we get

$$S_n^{(h)} \rightsquigarrow S^{(h)} = \sup_{s\in[0,1]} \int_{[0,1]^h} \{\mathbb{D}_C^{(h)}(s,\boldsymbol{u})\}^2\, \mathrm{d}C^{(h)}(\boldsymbol{u}). \tag{2.11}$$

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

2.1.3 Bootstrap and computation of approximate p-values

The null weak limit of $S_n^{(h)}$ in (2.11) is unfortunately untractable. Starting from Proposition 2.2 and adapting the approach of Bücher and Kojadinovic (2016a) and Bücher et al. (2014), we propose to base the computation of approximate p-values for $S_n^{(h)}$ on *multiplier* resampling versions of $\mathbb{C}_n^{(h)}$ in (2.6). For any $m \in \mathbb{N}$ and any $(s,t,\boldsymbol{u}) \in \Delta\times[0,1]^h$, let

$$\hat{\mathbb{C}}_n^{(h),[m]}(s,t,\boldsymbol{u}) = \hat{\mathbb{B}}_n^{(h),[m]}(s,t,\boldsymbol{u}) - \sum_{j=1}^h \dot{C}_{j,1:n}^{(h)}(\boldsymbol{u})\, \hat{\mathbb{B}}_n^{(h),[m]}(s,t,\boldsymbol{u}^{(j)}), \tag{2.12}$$

where

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

with $\boldsymbol{e}_j$ the j-th unit vector and

$$\hat{\mathbb{B}}_n^{(h),[m]}(s,t,\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \xi_{i,n}^{[m]} \left[\prod_{j=1}^{h} \mathbf{1}\{G_{1:n}(X_{i+j-1}) \leq u_j\} - C_{1:n}^{(h)}(\boldsymbol{u}) \right], \tag{2.13}$$

with $C_{1:n}^{(h)}$ and $G_{1:n}$ defined in (2.1) and (2.2), respectively. The sequences of random variables $(\xi_{i,n}^{[m]})_{i\in\mathbb{Z}}$, $m \in \mathbb{N}$, appearing in the expressions of the processes $\hat{\mathbb{B}}_n^{(h),[m]}$ in (2.13), $m \in \mathbb{N}$, are independent copies of what was called a *dependent multiplier sequence* in Bücher and Kojadinovic (2016a). 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, i.e., $\xi_{i,n}$ 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 somehow 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 (2016a, Section 5.2). In the rest of this work, we use the so-called *moving average approach* based on an initial independent and identically distributed (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

$$\xi_{i,n} = \sum_{j=1}^{\ell_n} \tilde{w}_{j,n} Z_{j+i-1}. \tag{2.14}$$

Then, as verified in Bücher and Kojadinovic (2016a, 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 test under consideration. In practice, we apply to the available univariate sequence $X_1, \dots, X_{n+h-1}$ the data-adaptive procedure proposed in Bücher and Kojadinovic (2016a, Section 5.1), 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 $X_1, \dots, X_{n+h-1}$.

Starting from (2.12) and having (2.5) in mind, multiplier resampling versions of $\mathbb{D}_n^{(h)}$ are then naturally given, for any $m \in \mathbb{N}$ and $(s, \boldsymbol{u}) \in [0,1]^{h+1}$, by

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

Corresponding multiplier resampling versions of the statistic $S_n^{(h)}$ in (2.3) are finally

$$\hat{S}_n^{(h),[m]} = \sup_{s\in[0,1]} \int_{[0,1]^h} \{\hat{\mathbb{D}}_n^{(h),[m]}(s, \boldsymbol{u})\}^2 \,\mathrm{d}C_{1:n}^{(h)}(\boldsymbol{u}),$$

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

The following proposition establishes the asymptotic validity of the multiplier resampling scheme under the null hypothesis of stationarity. It relies on the results of Bücher et al. (2014) and Bücher and Kojadinovic (2016a). It is proved in Appendix B.

Proposition 2.4. *Assume that $X_1, \dots, X_{n+h-1}$ are drawn from a strictly stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of continuous random variables whose strong mixing coefficients satisfy $\alpha_r^X = O(r^{-a})$ as $r \to \infty$ for some $a > 3 + 3h/2$, and $(\xi_{i,n}^{[1]})_{i\in\mathbb{Z}}, (\xi_{i,n}^{[2]})_{i\in\mathbb{Z}}, \dots$ are independent copies of a dependent multiplier sequence satisfying ($\mathcal{M}1$)–($\mathcal{M}3$) with $\ell_n = O(n^{1/2-\gamma})$ for some $0 < \gamma < 1/2$. Then, for any $M \in \mathbb{N}$,*

$$\big(\mathbb{C}_n^{(h)}, \hat{\mathbb{C}}_n^{(h),[1]}, \dots, \hat{\mathbb{C}}_n^{(h),[M]}\big) \rightsquigarrow \big(\mathbb{C}_C^{(h)}, \mathbb{C}_C^{(h),[1]}, \dots, \mathbb{C}_C^{(h),[M]}\big)$$

in $\{\ell^\infty(\Delta \times [0,1]^h)\}^{M+1}$, where $\mathbb{C}_C^{(h)}$ is defined in (2.9), and $\mathbb{C}_C^{(h),[1]}, \dots, \mathbb{C}_C^{(h),[M]}$ are independent copies of $\mathbb{C}_C^{(h)}$. As a consequence, for any $M \in \mathbb{N}$,

$$\big(\mathbb{D}_n^{(h)}, \hat{\mathbb{D}}_n^{(h),[1]}, \dots, \hat{\mathbb{D}}_n^{(h),[M]}\big) \rightsquigarrow \big(\mathbb{D}_C^{(h)}, \mathbb{D}_C^{(h),[1]}, \dots, \mathbb{D}_C^{(h),[M]}\big),$$

in $\{\ell^\infty([0,1]^{h+1})\}^{M+1}$, where $\mathbb{D}_C^{(h)}$ is defined in (2.10) and $\mathbb{D}_C^{(h),[1]}, \dots, \mathbb{D}_C^{(h),[M]}$ are independent copies of $\mathbb{D}_C^{(h)}$. Finally, for any $M \in \mathbb{N}$,

$$\big(S_n^{(h)}, \hat{S}_n^{(h),[1]}, \dots, \hat{S}_n^{(h),[M]}\big) \rightsquigarrow \big(S^{(h)}, S^{(h),[1]}, \dots, S^{(h),[M]}\big),$$

where $S^{(h)}$ is defined in (2.11) and $S^{(h),[1]}, \dots, S^{(h),[M]}$ are independent copies of $S^{(h)}$.

Note that, by Lemma 2.2 of Bücher and Kojadinovic (2017) and absolute continuity of $S^{(h)}$ (see Proposition 2.3 above), the last statement of Proposition 2.4 is equivalent to the following more classical formulation of bootstrap consistency:

$$d_K\big(\mathcal{L}(\hat{S}_n^{(h),[1]} \mid \boldsymbol{X}_n), \mathcal{L}(S_n^{(h)})\big) \equiv \sup_{x\in\mathbb{R}} |\mathrm{P}(\hat{S}_n^{(h),[1]} \leq x \mid \boldsymbol{X}_n) - \mathrm{P}(S_n^{(h)} \leq x)| \xrightarrow{\mathrm{P}} 0.$$

Here, d_K is the Kolmogorov distance between two probability laws on the real line, $\boldsymbol{X}_n = (X_1, \dots, X_{n+h-1})$ denotes the available data, and $\mathcal{L}(\hat{S}_n^{(h),[1]} \mid \boldsymbol{X}_n)$ and $\mathcal{L}(S_n^{(h)})$ denote (a regular versions of) the conditional law of $S_n^{(h),[1]}$ given $\boldsymbol{X}_n$ and the law of $S_n^{(h)}$, respectively. Furthermore, Lemma 4.1 in Bücher and Kojadinovic (2017) ensures that the test based on $S_n^{(h)}$ with approximate p-value $p_{n,M}(S_n^{(h)}) = M^{-1}\sum_{m=1}^M \mathbf{1}\big(\hat{S}_n^{(h),[m]} \geq S_n^{(h)}\big)$ holds its level asymptotically under the null hypothesis of stationarity as n and M tend to the infinity. By Corollary 4.2 in the same reference, this implies that $p_{n,M_n}(S_n^{(h)}) \rightsquigarrow \mathrm{Uniform}(0,1)$ when $n \to \infty$, for any sequence $M_n \to \infty$.

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

We propose to combine the previous test with a test particularity sensitive to departures from $H_0^{(1)}$ 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 $X_1, \dots, X_n$, thereby ignoring the remaining $h-1$ ones. The test statistic can then be written as

$$T_n = \sup_{s \in [0,1]} \int_{\mathbb{R}} \{\mathbb{E}_n(s,x)\}^2 \, \mathrm{d}G_{1:n}(x), \tag{2.15}$$

where

$$\mathbb{E}_n(s,x) = \sqrt{n}\lambda_n(0,s)\lambda_n(s,1)\left\{G_{1:\lfloor ns \rfloor}(x) - G_{\lfloor ns \rfloor+1:n}(x)\right\}, \qquad (s,x) \in [0,1] \times \mathbb{R}, \tag{2.16}$$

and, for any $1 \le k \le l \le n$, $G_{k:l}$, $1 \le k \le l \le n$, is defined as in (2.2) 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 (2.16) 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

$$\mathbb{G}_n(s,x) = \sqrt{n}\,\lambda_n(0,s)\,\{G_{1:\lfloor ns \rfloor}(x) - G(x)\}, \qquad (s,x) \in [0,1] \times \mathbb{R}. \tag{2.17}$$

The following result, proved in Appendix B and providing the null weak limit of T_n in (2.15), is partly an immediate consequence of Theorem 1 of Bücher (2015) and of the continuous mapping theorem.

Proposition 2.5. *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, $\mathbb{G}_n \leadsto \mathbb{G}$ 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, $\mathbb{E}_n \leadsto \mathbb{E}$ in $\ell^\infty([0,1] \times \mathbb{R})$, where

$$\mathbb{E}(s,x) = \mathbb{G}(s,x) - s\mathbb{G}(1,x), \qquad (s,x) \in [0,1] \times \mathbb{R}, \tag{2.18}$$

and $T_n \leadsto T$ with

$$T = \sup_{s \in [0,1]} \int_{\mathbb{R}} \{\mathbb{E}(s,x)\}^2 \, \mathrm{d}G(x). \tag{2.19}$$

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

Following Gombay and Horváth (1999), Holmes et al. (2013) and Bücher and Kojadinovic (2016a), we shall compute approximate p-values for T_n using multiplier resampling versions of $\mathbb{G}_n$ in (2.17). 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

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

An approximate p-value for T_n will then be computed as $p_{n,M}(T_n) = M^{-1} \sum_{m=1}^{M} \mathbf{1}\big(\hat{T}_n^{[m]} \geq T_n\big)$ 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 $S_n^{(h)}$ presented in the previous section. The result is a direct consequence of Corollary 2.2 in Bücher and Kojadinovic (2016a); see also Proposition A.1 in the next section. In particular, $p_{n,M_n}(T_n) \rightsquigarrow \text{Uniform}(0,1)$ for any sequence $(M_n)_n$ that tends to the infinity as $n \to \infty$.

2.3 Combining the two tests

To be able to combine the two tests based on $S_n^{(h)}$ and T_n, we shall carry them out using the same M dependent multiplier sequences. Since a moving average approach is used to generate such sequences, it follows from (2.14) that it is sufficient to impose that the same M initial normal sequences are used for both tests. To obtain a global p-value, we propose appropriate extensions to dependent tests of 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 hypotheses $H^{(1)}, \dots, H^{(r)}$ with corresponding positive weights $w_1, \dots, w_r$ that sum up to one and quantify the importance of each test in the combination, the latter method consists of computing, up to a rescaling term, the global statistic

$$\psi_S(p_1, \dots, p_r) = \sum_{i=1}^{r} w_i \Phi^{-1}(1 - p_i), \tag{2.20}$$

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

$$\psi_F(p_1, \dots, p_r) = -2 \sum_{i=1}^{r} w_i \log(p_i). \tag{2.21}$$

In the rest of this section, we restrict ourselves to the combination of the two tests based on $S_n^{(h)}$ in (2.3) and T_n in (2.15) but it is important to keep in mind that the proposed methodology straightforwardly extends to more than two tests. Let ψ be a continuous function from $(0,1)^2$ to $\mathbb{R}$ that is decreasing in both of its arguments. The proposed combination procedure is as follows:

1. Let $\hat{S}_n^{(h),[0]} = S_n^{(h)}$ and $\hat{T}_n^{[0]} = T_n$ be the statistics computed from $X_1, \dots, X_{n+h-1}$.
2. Estimate ℓ_n from $X_1, \dots, X_{n+h-1}$ as explained in Section 2.1.3.
3. Given a large integer M and the value of ℓ_n obtained in Step 2, generate M independent copies of the same dependent multiplier sequence using (2.14) and compute the corresponding multiplier replicates $(\hat{S}_n^{(h),[1]}, \hat{T}_n^{[1]}), \dots, (\hat{S}_n^{(h),[M]}, \hat{T}_n^{[M]})$ of $(S_n^{(h)}, T_n)$.
4. Then, for all $j \in \{0, 1, \dots, M\}$, compute

$$p_{n,M}(\hat{S}_n^{(h),[j]}) = \frac{1}{M+1} \left\{ \frac{1}{2} + \sum_{i=1}^{M} \mathbf{1}\left(\hat{S}_n^{(h),[i]} \geq \hat{S}_n^{(h),[j]} \right) \right\},$$

and

$$p_{n,M}(\hat{T}_n^{[j]}) = \frac{1}{M+1} \left\{ \frac{1}{2} + \sum_{i=1}^{M} \mathbf{1}\left(\hat{T}_n^{[i]} \geq \hat{T}_n^{[j]} \right) \right\}.$$

5. Next, for all $i \in \{0, 1, \dots, M\}$, compute
$$W_{n,M}^{[i]} = \psi\{p_{n,M}(\hat{S}_n^{(h),[i]}), p_{n,M}(\hat{T}_n^{[i]})\}.$$
6. The global statistic is $W_{n,M}^{[0]}$ and the corresponding approximate p-value is given by
$$p_{n,M}(W_{n,M}^{[0]}) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}\left(W_{n,M}^{[i]} \geq W_{n,M}^{[0]}\right). \tag{2.22}$$

Note that the quantities $p_{n,M}(\hat{S}_n^{(h),[j]})$ and $p_{n,M}(\hat{T}_n^{[j]})$ in Step 4 can be regarded as approximate p-values for the "statistic values" $S_n^{(h),[j]}$ and $T_n^{[j]}$, respectively. The offset by $1/2$ and the division by $M+1$ instead of M in the formulas is carried out to ensure that $p_{n,M}(\hat{S}_n^{(h),[j]})$ and $p_{n,M}(\hat{T}_n^{[j]})$ belong to the interval $(0,1)$ so that Step 5 is well-defined.

The next result, proved in Appendix B, states, in three slightly different ways, the asymptotic validity of the combined test under the hypothesis of stationarity and the natural assumption that $M = M_n \to \infty$ as $n \to \infty$. Before proceeding, recall that $\boldsymbol{X}_n = (X_1, \dots, X_{n+h-1})$ denotes the available data and note that the statistic $W_{n,M_n}^{[0]}$ of the combined test is a Monte Carlo approximation of the unobservable statistic $W_n = \psi\{\mathrm{P}(\hat{S}_n^{(h),[1]} \geq S_n^{(h)} \mid \boldsymbol{X}_n), \mathrm{P}(\hat{T}_n^{[i]} \geq T_n \mid \boldsymbol{X}_n)\}$.

Proposition 2.6. *Let $M = M_n \to \infty$ as $n \to \infty$. Under the conditions of Proposition 2.4, for any $N \in \mathbb{N}$,*
$$(W_{n,M_n}^{[0]}, W_{n,M_n}^{[1]}, \dots, W_{n,M_n}^{[N]}) \rightsquigarrow (W, W^{[1]}, \dots, W^{[N]}), \tag{2.23}$$
where $W = \psi\{\bar{F}_S(S^{(h)}), \bar{F}_T(T)\}$ is the weak limit of W_n, with $\bar{F}_S(x) = \mathrm{P}(S^{(h)} \geq x)$, $x \in \mathbb{R}$, and $\bar{F}_T(x) = \mathrm{P}(T \geq x)$, $x \in \mathbb{R}$, and $W^{[1]}, \dots, W^{[N]}$ are independent copies of W. If, additionally, ψ is chosen in such a way that the random variable W is continuous, then
$$d_K\big(\mathcal{L}(W_{n,M_n}^{[1]} \mid \boldsymbol{X}_n), \mathcal{L}(W_n)\big) = \sup_{x\in\mathbb{R}} |\mathrm{P}(W_{n,M_n}^{[1]} \leq x \mid \boldsymbol{X}_n) - \mathrm{P}(W_n \leq x)| \xrightarrow{\mathrm{P}} 0, \tag{2.24}$$
and
$$\sup_{x\in\mathbb{R}} \left| \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}(W_{n,M_n}^{[i]} \leq x) - \mathrm{P}(W_n \leq x) \right| \xrightarrow{\mathrm{P}} 0. \tag{2.25}$$
As a consequence, $p_{n,M_n}(W_{n,M_n}^{[0]}) \rightsquigarrow$ Uniform$(0,1)$, where $p_{n,M_n}(W_{n,M_n}^{[0]})$ is defined in (2.22).

It is worthwhile mentioning that the unconditional bootstrap consistency statement in (2.23) does not require W to be a continuous random variable. Proving the latter might actually be quite complicated: unlike $S^{(h)}$ in (2.11) and T in (2.19), 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 $(S^{(h)}, T)$ 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 be continuous 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 T_n or $S_n^{(h)}$ (for h suitably chosen) converges to infinity in probability, while their bootstrap versions are $O_{\mathrm{P}}(1)$. As a consequence, provided $\psi(p_1, p_2) \to \infty$ when $\min(p_1, p_2) \to 0$, one obtains that $W_{n,M_n}^{[0]}$ converges to infinity in probability as well, while the bootstrap replicates $W_{n,M_n}^{[i]}, i \geq 1$, are $O_{\mathrm{P}}(1)$. These two facts imply consistency of the test based on $W_{n,M_n}^{[0]}$. Further details are omitted for the sake of brevity.

3 On the choice of the embedding dimension h

The methodology described in the previous section 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, we will make a practical suggestion at the end of this section.

Let us start by considering arguments in favour 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 (multivariate) 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, has not yet a general solution. 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 $H_{0,c}^{(h+1)}$ seem to be.

The previous paragraph suggests to choose h as large as possible, even if the marginal gain of an increase of h becomes smaller for larger and larger h. At the opposite, there are also good reasons for choosing h rather small. Indeed, for many sensible models, the power of the test based on $S_n^{(h)}$ in (2.3) is a decreasing function of h, at least from some small value onwards. This observation will for instance be one of the results of our simulation study in Section 5, but it can also be supported by more theoretical arguments. Indeed, consider for instance 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 obtain that

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

where $C_{s^*}^{(h)} = s^* C_1^{(h)} + (1-s^*)C_2^{(h)}$. In other words, the dominating term in an asymptotic expansion of $S_n^{(h)}$ converges to infinity at rate n, with scaling factor κ_h depending on h. Since the bootstrap replicates of $S_n^{(h)}$ are $O_{\mathrm{P}}(1)$ for any h (which can be proved along the lines of Theorem 3 (ii) in Holmes et al., 2013), we 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 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.

Additionally, 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, 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 by 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 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, e.g., 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 (2016b) and proceeding as in Section 2, 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

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

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

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

where

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

and $\mathbb{U}_n(s) = 0$ 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 $\phi(z,z') = m(z,z') = z$, $z,z' \in \mathbb{R}$, the statistic $R_{\phi,n}^{(q)} = R_{m,n}^{(1)}$ is (asymptotically equivalent to) the classical CUSUM statistic that is particularly sensitive to changes in the 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 $R_{v,n}^{(1)}$ particularly sensitive to changes in the variance of the observations. For $q \in \{2,\dots,h\}$, setting $\phi(\boldsymbol{z},\boldsymbol{z}') = a(\boldsymbol{z},\boldsymbol{z}') = (z_1 - z_1')(z_2 - z_2')/2$, $\boldsymbol{z},\boldsymbol{z}' \in \mathbb{R}^2$, results in the CUSUM statistic $R_{a,n}^{(q)}$ sensitive to changes in the autocovariance at lag $q-1$.

From Bücher and Kojadinovic (2016b), CUSUM tests based on $R_{m,n}^{(1)}$, $R_{v,n}^{(1)}$ and $R_{a,n}^{(q)}$, $q \in \{2,\dots,h\}$, 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 Section 2.3. Specifically, for the generic test based on $R_{\phi,n}^{(q)}$, 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

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

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

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

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

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

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

Let ψ be a continuous function from $(0,1)^{h+1}$ to $\mathbb{R}$ that is decreasing in each of its arguments. The analogue of the testing procedure of Section 2.3 for combining the tests based on $R_{m,n}^{(1)}$, $R_{v,n}^{(1)}$ and $R_{a,n}^{(q)}$, $q \in \{2,\dots,h\}$, is as follows:

1. Let $\hat{R}_{\phi,n}^{(q),[0]} = R_{\phi,n}^{(q)}$ for $q = 1$ and $\phi \in \{m, v\}$, and for $q \in \{2,\dots,h\}$ and $\phi = a$.
2. Estimate ℓ_n from $X_1,\dots,X_n$ as explained in Bücher and Kojadinovic (2016b, Section 2.4) for $\phi = m$.
3. Given a large integer M and the value of ℓ_n obtained in Step 2, generate M independent copies of the same dependent multiplier sequence using (2.14) and compute the corresponding multiplier replicates $\hat{R}_{\phi,n}^{(q),[1]},\dots,\hat{R}_{\phi,n}^{(q),[M]}$ for $q = 1$ and $\phi \in \{m, v\}$, and for $q \in \{2,\dots,h\}$ and $\phi = a$.

4. Then, for all $j \in \{0, 1, \dots, M\}$, compute
$$p_{n,M}(\hat{R}^{(q),[j]}_{\phi,n}) = \frac{1}{M+1} \left\{ \frac{1}{2} + \sum_{i=1}^{M} \mathbf{1} \left(\hat{R}^{(q),[i]}_{\phi,n} \geq \hat{R}^{(q),[j]}_{\phi,n} \right) \right\},$$
for $q = 1$ and $\phi \in \{m, v\}$, and for $q \in \{2, \dots, h\}$ and $\phi = a$.

5. Next, for all $i \in \{0, 1, \dots, M\}$, compute
$$W^{[i]}_{n,M} = \psi\{p_{n,M}(\hat{R}^{(1),[i]}_{m,n}), p_{n,M}(\hat{R}^{(1),[i]}_{v,n}), p_{n,M}(\hat{R}^{(2),[i]}_{a,n}), \dots, p_{n,M}(\hat{R}^{(h),[i]}_{a,n})\}.$$

6. The global statistic is $W^{[0]}_{n,M}$ and the corresponding approximate p-value is given by
$$p_{n,M}(W^{[0]}_{n,M}) = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1} \left(W^{[i]}_{n,M} \geq W^{[0]}_{n,M} \right). \tag{4.4}$$

Conditions under which the previous testing procedure is asymptotically valid under the null hypothesis of stationarity can be obtained by starting from Proposition 2.5 in Bücher and Kojadinovic (2016b) and by proceeding as in the proofs of the results stated in Section 2.3. For the sake of simplicity, the conditions in the following proposition 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
$$\beta^Z_r = \sup_{p\in\mathbb{Z}} \mathrm{E} \sup_{A\in\mathcal{F}^{\infty}_{p+r}} |\mathrm{P}(A \mid \mathcal{F}^{p}_{-\infty}) - \mathrm{P}(A)|, \qquad r \in \mathbb{N}, r > 0,$$
where $\mathcal{F}^b_a$ is defined above (2.8). 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 $\alpha^Z_r \leq \beta^Z_r$, absolute regularity implies strong mixing.

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 $\delta > 0$. Also, let $(\xi^{[1]}_{i,n})_{i\in\mathbb{Z}}, (\xi^{[2]}_{i,n})_{i\in\mathbb{Z}}, \dots$ be independent copies of a dependent multiplier sequence satisfying $(\mathcal{M}1)$–$(\mathcal{M}3)$ with $\ell_n = O(n^{1/2-\gamma})$ for some $1/(6+2\delta) < \gamma < 1/2$. Then, if $\beta^X_r = O(r^{-b})$ for some $b > 2(4+\delta)/\delta$ as $r \to \infty$, $p_{n,M_n}(W^{[0]}_{n,M_n}) \rightsquigarrow$ Uniform$(0,1)$ for any $M_n \to \infty$, where $p_{n,M_n}(W^{[0]}_{n,M_n})$ is defined by (4.4).*

5 Monte Carlo experiments

Various simulations were carried out in order to study the finite-sample performance of the tests that have been proposed in Sections 2 and 4. After describing the main data generating models, we investigate the behavior of some competitor tests available in R, study how well the proposed tests hold their level and finally assess their power under various alternatives, some belonging to the locally stationary process literature, others more in line with the change-point detection literature.

5.1 Data generating processes

The following ten strictly stationary models were used to generate observations under the null hypothesis of stationarity. Either standard normal or standardized Student t with 4 degrees of freedom innovations were considered (standardization refers to the fact that the Student t with 4 degrees of freedom distribution was rescaled to have variance one). The first seven models were considered in Nason (2013) (and are denoted by S1–S7 therein):

N1 - i.i.d. observations from the innovation distribution.

N2 - AR(1) model with parameter 0.9.

N3 - AR(1) model with parameter -0.9.

N4 - MA(1) model with parameter 0.8.

N5 - MA(1) model with parameter -0.8.

N6 - ARMA(1, 0, 2) with the AR coefficient -0.4, and the MA coefficients $(-0.8, 0.4)$.

N7 - AR(2) with parameters 1.385929 and -0.9604. This process is stationary, but close to the "unit root": a "rough" stochastic process with spectral peak near $\pi/4$.

N8 - GARCH(1,1) model with parameters $(\omega, \beta, \alpha) = (0.012, 0.919, 0.072)$. The latter values were estimated by Jondeau et al. (2007) from SP500 daily log-returns.

N9 - the exponential autoregressive model considered in Auestad and Tjøstheim (1990) whose generating equation is

$$X_t = \{0.8 - 1.1\exp(-50X_{t-1}^2)\}X_{t-1} + 0.1\varepsilon_t.$$

N10 - the nonlinear autoregressive model used in Paparoditis and Politis (2001, Section 3.3) whose generating equation is

$$X_t = 0.6\sin(X_{t-1}) + \varepsilon_t.$$

For all these models, a burn-in period of 100 observations was used.

To simulate observations under the alternative hypothesis of non-stationarity, models connected to the literature on locally stationary processes were considered first. The first four are taken from Dette et al. (2011) and were used to generate univariate series $X_{1,n}, \dots, X_{n,n}$ of length $n \in \{128, 256, 512\}$ by means of the following equations:

A1 - $X_{t,n} = 1.1\cos\{1.5 - \cos(4\pi t/n)\}\varepsilon_{t-1} + \varepsilon_t$,

A2 - $X_{t,n} = 0.6\sin(4\pi t/n)X_{t-1,n} + \varepsilon_t$,

A3 - $X_{t,n} = (0.5X_{t-1,n} + \varepsilon_t)\mathbf{1}(t \in \{1, \dots, n/4\} \cup \{3n/4+1, \dots, n\}) + (-0.5X_{t-1,n} + \varepsilon_t)\mathbf{1}(t \in \{n/4+1, \dots, 3n/4\})$,

A4 - $X_{t,n} = (-0.5X_{t-1,n} + \varepsilon_t)\mathbf{1}(t \in \{1, \dots, n/2\} \cup \{n/2+n/64+1, \dots, n\}) + 4\varepsilon_t\mathbf{1}(t \in \{n/2+1, \dots, n/2+n/64\})$,

where $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$ are i.i.d. standard normal and with the convention that $X_{0,n} = 0$. The next four models under the alternative were considered in Nason (2013) (and are denoted P1–P4 therein), the last three ones being locally stationary wavelet (LSW) processes (see, e.g., Equation (1) in Nason, 2013):

A5 - A time-varying AR model $X_t = \alpha_t X_{t-1} + \varepsilon_t$ with i.i.d. standard normal innovations and an AR parameter evolving linearly from 0.9 to -0.9 over the n observations.

A6 - A LSW process based on Haar wavelets with spectrum $S_j(z) = 0$ for $j > 1$ and $S_1(z) = 1/4 - (z - 1/2)^2$ for $z \in (0, 1)$. This process is a time-varying moving average process.

A7 - A LSW process based on Haar wavelets with spectrum $S_j(z) = 0$ for $j > 2$, $S_1(z)$ as for A6 and $S_2(z) = S_1(z + 1/2)$ using periodic boundaries (for the construction of the spectrum only).

A8 - A LSW process based on Haar wavelets with spectrum $S_j(z) = 0$ for $j = 2$ and $j > 4$. Moreover, $S_1(z) = \exp\{-4(z - 1/2)^2\}$, $S_3(z) = S_1(z - 1/4)$ and $S_4(z) = S_1(z + 1/4)$, again assuming periodic boundaries.

Models A1–A8 considered thus far are connected to the literature on locally stationary processes. In a second set of experiments, we focused on models that are more in line with the change-point detection literature:

A9 - An AR(1) model with one break: the $n/2$ first observations are i.i.d. from the innovation distribution (standard normal or standardized t_4) and the $n/2$ last observations are from an AR(1) model with parameter $\beta \in \{-0.8, -0.4, 0, 0.4, 0.8\}$.

A10 - An AR(2) model with one break: the $n/2$ first observations are i.i.d. from the innovation distribution (standard normal or standardized t_4) and the $n/2$ last observations are from an AR(2) model with parameter $(0, \beta)$ with $\beta \in \{-0.8, -0.4, 0, 0.4, 0.8\}$.

Note that both the contemporary distribution and the serial dependence are changing under these scenarios (unless $\beta = 0$). Also note that there is no relationship between X_t and X_{t-1} for Model A10, that is, $H_{0,c}^{(2)}$ in (1.4) is met with $C^{(2)}$ the bivariate independence copula.

Finally, we considered two simple models under the alternative for which the contemporary distribution remains unchanged:

A11 - An AR(1) model with a break affecting the innovation variance: 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 $\beta \in \{0, 0.4, 0.8\}$ and centered normal innovations with variance $(1 - \beta^2)$. The contemporary distribution is thus the standard normal.

A12 - A max-autoregressive model with one break: the $n/2$ first observations are i.i.d. standard Fréchet, and the last $n/2$ observations follow the recursion

$$X_t = \max\{\beta X_{t-1}, (1 - \beta) Z_t\},$$

where $\beta \in \{0, 0.4, 0.8\}$ and the Z_t are i.i.d standard Fréchet. The contemporary distribution is standard Fréchet regardless of the choice of β, see, e.g., Example 10.3 in Beirlant et al. (2004).

5.2 Some competitors to our tests

As mentioned in the introduction, many tests of stationarity have been proposed in the literature. Unfortunately, only a few of them seem to have been implemented in statistical software. In this section, we focus on the tests of Priestley and Subba Rao (1969), Nason (2013) and Cardinali and Nason (2013) that have been 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 include the test of Cardinali and Nason (2016) (implemented in the R package `BootWPTOS`) in our simulations because we were not able to understand how to initialize the arguments of the corresponding R function.

The rejection percentages of the three aforementioned tests were estimated for Models N1–N10 generating observations under the null. These tests were carried out at the 5% significance level and the empirical levels were estimated from 1000 samples. As one can see in Table 1, all these tests turn out to be too liberal in at least one scenario. Overall, the empirical levels are even higher when heavy tailed innovations are used instead of standard normal ones.

5.3 Empirical levels of the proposed tests

To estimate the levels of the proposed tests, we considered the same setting as in the previous section. As we continue, we shall refer to the tests to be combined as the *component* tests as

Table 1: Percentage of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n \in \{128, 256, 512\}$ generated from Models N1–N10, first with $N(0,1)$ innovations and then with standardized t_4 innovations. The column PSR.T corresponds to the test of Priestley and Subba Rao (1969) implemented in the R package `fractal`, the column hwtos2 corresponds to the test of Nason (2013) implemented in the R package `locits` and the column BTOS corresponds to the test of Cardinali and Nason (2013) implemented in the R package `costat`.

		$N(0,1)$ innovations			Standardized t_4 innovations		
Model	n	PSR.T	hwtos2	BTOS	PSR.T	hwtos2	BTOS
N1	128	7.0	0.4	0.0	30.6	4.2	5.8
	256	5.7	3.3	0.0	47.3	15.1	8.5
	512	5.9	3.0	0.1	62.1	19.0	13.5
N2	128	46.3	1.0	0.0	66.6	2.4	0.0
	256	22.0	3.6	0.0	59.5	12.5	0.0
	512	11.1	4.6	0.0	65.5	13.6	0.0
N3	128	6.8	9.7	21.1	31.1	12.5	32.2
	256	6.7	15.8	35.8	46.0	22.8	41.4
	512	5.7	17.7	38.5	63.3	29.1	52.6
N4	128	7.2	2.0	0.0	34.8	3.5	1.0
	256	7.4	4.7	0.0	46.8	18.8	2.4
	512	6.3	3.9	0.0	61.1	15.8	2.6
N5	128	12.7	0.0	0.4	38.1	0.1	6.6
	256	9.1	0.0	0.5	48.3	7.4	11.6
	512	6.3	0.3	0.8	64.1	6.7	16.7
N6	128	24.3	0.2	2.0	44.0	0.2	12.9
	256	8.6	0.7	2.4	46.8	6.5	16.6
	512	7.1	0.2	3.2	63.7	4.9	21.1
N7	128	62.8	1.1	0.5	63.8	1.3	0.6
	256	63.1	7.1	5.0	73.5	9.1	8.3
	512	49.2	20.6	17.5	78.6	29.0	22.5
N8	128	20.1	1.9	2.5	58.1	6.0	17.4
	256	38.5	8.9	4.1	81.0	30.1	33.8
	512	56.5	11.5	6.5	94.8	44.3	47.9
N9	128	34.8	5.9	0.0	68.9	13.6	0.4
	256	25.8	16.6	0.2	71.3	34.8	1.6
	512	17.9	22.8	0.5	75.6	41.9	3.1
N10	128	5.2	0.4	0.0	19.8	1.6	3.7
	256	4.2	2.0	0.0	36.1	12.2	5.3
	512	4.6	2.7	0.1	61.1	12.3	8.7

opposed to the *combined* tests. We started with the component tests described in Sections 2 and 4, and then considered various combinations of them, based on the weighted version of Stouffer's method and a weighted version of Fisher's method. As explained previously, the former (resp. latter) consists of using ψ_S in (2.20) (resp. ψ_F in (2.21)) as the function ψ in Sections 2 and 4. As Stouffer's method sometimes gave inflated levels, for the sake of brevity, we only report the results for Fisher's method in this section.

As previously, all the tests were carried out at the 5% significance level and the empirical levels were estimated from 1000 samples generated from Models N1–N10. The values 128, 256 and 512 were considered for the sample size n and the embedding dimension h was taken to be in the set $\{2, 3, 4, 8\}$. To save space in the tables providing the results, each component test is abbreviated by a single letter:

- d for the d.f. test based on T_n in (2.15),
- c for the empirical autocopula test at lag $h-1$ based on $S_n^{(h)}$ in (2.3) (the value of h will always be clear from the context),
- m for the sample mean test based on $R_{n,m}^{(1)}$ defined generically by (4.3),
- v for the variance test based on $R_{n,v}^{(1)}$ defined generically by (4.3), and
- a for the autocovariance test at lag $q-1$ based on $R_{n,a}^{(q)}$, $q \in \{2, \dots, h\}$, 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 $1/\{3(h-1)\}$,
- 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 T_n in (2.15) is combined with $S_n^{(2)}$ in (2.3) and $\tilde{S}_n^{(3)}, \dots, \tilde{S}_n^{(h)}$, where the latter are the analogues of $S_n^{(2)}$ but for lags $2, \dots, h-1$ (that is, 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.

The empirical levels of the tests are reported in Tables 2 and 3 for $h \in \{2, 3, 4, 8\}$ (for $h = 8$, to save computing time, only the tests c and dc were carried out). As one can see, the rank-based tests d, c, dc and dcp of Section 2 never turned out, overall, to be too liberal (unlike their competitors considered in Section 5.2 – see Table 1). Their analogues of Section 4 focusing on second-order characteristics behave reasonably well except for Model N8. The latter is due to the fact that the test v is way too liberal for Model N8 as can be seen from Table 3, a probable consequence of the conditional heteroskedasticity of the model. For fixed n, as h increases, the empirical autocopula test c (and thus the combined test dc) can be seen to be more and more conservative, as already mentioned in Section 3. The latter clearly appears by considering the last vertical blocks of Tables 2 and 3 corresponding to $h = 8$. Nonetheless, the rejection percentages therein hint at the fact that the empirical levels, as expected theoretically, should improve as n increases further.

5.4 Empirical powers

The empirical powers of the proposed tests were estimated under Models A1–A12 from 1000 samples of size $n \in \{128, 256, 512\}$ for $h \in \{2, 3, 4, 8\}$ (again, for $h = 8$, only the tests c and dc were carried out). For Models A1–A8 (those that are connected to the literature on locally stationary processes), the rejection percentages of the null hypothesis of stationarity are reported in Table 4. As one can see, for $h \in \{2, 3, 4\}$, the rank-based combined tests of Section 2.3 (dc and dcp) almost always seem to be more powerful than the combined tests of second-order stationarity that have been considered in Section 4 (va and mva). Furthermore, the tests focusing on the contemporary distribution (d, m and v) hardly have any power overall, suggesting that the distribution of X_t does not change (too) much for the models under consideration (note in passing the very disappointing behavior of the test m for Models A6–A8). The latter explains why the test c is more powerful than the combined tests dc and dcp, and why the test a is almost always more powerful than va and mva for $h = 2$. Finally, note that, except for A8, the power of all the tests focusing on serial dependence decreases, overall, as h increases (see also the discussion in Section 3). At least for Models A1–A4, the latter is a consequence of the fact that the serial dependence is completely determined by the distribution of (X_t, X_{t-1}).

The results of Table 4 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) and in Section 5.2 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 Models A1–A4 (results for Models A5–A8 are not available in the latter reference). 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).

While Models A1–A8 considered thus far are connected to the literature on locally stationary processes, the remaining Models A9–A12 are more in line with the change-point literature. For the latter, all our tests (except m) turn out to display substantially more power. This should not come as a surprise given that the tests are based on the CUSUM approach and are hence designed to detect alternatives involving one single break.

Table 5 reports the empirical powers of the proposed tests for Model A9. Recall that both the contemporary distribution and the serial dependence is changing under this scenario (unless $\beta = 0$). As one can see, even in this setting that should possibly be favorable to the tests focusing on second-order stationarity, the rank-based tests involving test c appear more powerful, overall, than those involving test a. Furthermore, with a few exceptions, the test c is always at least slightly more powerful than the combined test dc. As expected given the data generating model and in line with the discussion of Section 3, the increase of h leads to a decrease in the power of c and dc. In addition, for $h \in \{3, 4\}$, dcp is more powerful than dc, which can be explained by the fact that the serial dependence in the data generating model is solely of a bivariate nature.

The rejection percentages for Model A10 are reported in Table 6. As expected, the empirical powers of tests c and a are very low for $h = 2$ since there is no relationship between X_t and X_{t-1}. The tests focusing on the contemporary distribution are more powerful, in particular the test v. Consequently, the combined tests at lag 1 involving v do display some power. For $h \in \{3, 4\}$, the two most powerful tests are dcp and va. The fact that dcp is more powerful than dc can again be explained by the bivariate nature of the serial dependence.

Finally, the rejection percentages for Models A11 and A12 are given in Table 7. The columns c2 and c3 report the results for the bivariate analogues of the tests based on $S_n^{(2)}$ defined by (2.3) for lags 2 and 3 (these tests arise in the combined test dcp). To save computing time, we did not include the tests of second-order stationarity as these were found less powerful, overall, in the previous experiments (for Models A12, moments do not exist, whence an application would not even be meaningful). Comparing the results for Model A11 with those of Table 5 for the same values of h reveals, as expected, a higher power of the test c. In addition, the test c for lag 1 is more powerful than the test c2, which, in turn, is more powerful than the test c3, a consequence of the data generating models. Finally, the fact the test d displays some power for $\beta = 0.8$ seems to be only a consequence of the sample sizes under consideration and the very strong serial dependence in the second half of the observations.

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 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 2 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 three 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 multivariate time series. As multivariate 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 8. 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. From the second horizontal block of Table 8, 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 test of stationarity proposed in Section 2 was never observed to be too liberal for the rather typical sample sizes considered in this work. As discussed in Section 3 and as empirically confirmed by the experiments of Section 5, 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.

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 Kirch and Muhsal (2016): 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 (2.4) (if the focus is on serial dependence) or on the process in (2.16) (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 multivariate 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, it will then be helpful in terms of finite sample power properties to split the hypothesis $H_0^{(h)}$ in (1.2) into suitable sub-hypotheses. For $A \subset \{1, \dots, d\}$ and $B \subset \{0, \dots, h-1\}$, let

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

Letting $\bar{d} = \{1, \dots, d\}$ and $\bar{h} = \{0, \dots, h-1\}$, Sklar's theorem suggests the decomposition $H_0^{(h)} = H_0^{(1)}(\{1\}) \cap \dots \cap H_0^{(1)}(\{d\}) \cap H_{0,c}^{(h)}(\bar{d}, \bar{h})$. However, preliminary numerical experiments indicate that a straightforward extension of the approach proposed in Section 2.3 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 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 multivariate case. As an imperfect alternative, one might for instance consider the following hypothesis

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

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, like for instance $\bigcap_{i\neq j\in\bar{d}} H_{0,c}^{(h)}(\{i,j\}, \{0,1\})$. The amount of potential adaptations appears to be very large, whence a further investigation, in particular from a finite-sample point-of-view, is beyond the scope of this paper.

A Auxiliary results

The proof of Proposition 2.6 relies on the following result, establishing the joint unconditional weak convergence of the two statistics to be combined along with their dependent multiplier replicates. The proof is given in Section B.

Proposition A.1. *Under the conditions of Proposition 2.4, for any $M \in \mathbb{N}$,*

$$\big((\mathbb{D}_n^{(h)}, \mathbb{E}_n), (\hat{\mathbb{D}}_n^{(h),[1]}, \hat{\mathbb{E}}_n^{[1]}), \dots, (\hat{\mathbb{D}}_n^{(h),[M]}, \hat{\mathbb{E}}_n^{[M]})\big) \rightsquigarrow \big((\mathbb{D}_C^{(h)}, \mathbb{E}), (\mathbb{D}_C^{(h),[1]}, \mathbb{E}^{[1]}), \dots, (\mathbb{D}_C^{(h),[M]}, \mathbb{E}^{[M]})\big)$$

in $\{\ell^\infty([0,1]\times\mathbb{R})\}^{2(M+1)}$, where $\mathbb{D}_C^{(h)}$ and $\mathbb{E}$ are defined in (2.10) and (2.18), respectively, and $(\mathbb{D}_C^{(h),[1]},\mathbb{E}^{[1]}),\dots,(\mathbb{D}_C^{(h),[M]},\mathbb{E}^{[M]})$ 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 $(\hat{\mathbb{D}}_n^{(h),[m]},\hat{\mathbb{E}}_n^{[m]})$ can be considered to have the same joint law as $(\mathbb{D}_C^{(h)},\mathbb{E})$ asymptotically. As a consequence, for any $M\in\mathbb{N}$,

$$\big((S_n^{(h)},T_n),(\hat{S}_n^{(h),[1]},\hat{T}_n^{[1]}),\dots,(\hat{S}_n^{(h),[M]},\hat{T}_n^{[M]})\big)\rightsquigarrow\big((S^{(h)},T),(S^{(h),[1]},T^{[1]}),\dots,(S^{(h),[M]},T^{[M]})\big),$$

where $S^{(h)}$ and T are defined in (2.11) and (2.19), respectively, and where the random vectors $(S^{(h),[1]},T^{[1]}),\dots,(S^{(h),[M]},T^{[M]})$ are independent copies of $(S^{(h)},T)$.

B Proofs

Proof of Proposition 2.2. 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 $\alpha_r^{\boldsymbol{Y}}=\alpha^X_{(r-h+1)\vee 0}$, $r\in\mathbb{N}$, where $\vee$ is the maximum operator. ■

Proof of Proposition 2.3. The assertions concerning weak convergence are simple consequences of the continuous mapping theorem and Proposition 2.2. It remains to show that $\mathcal{L}(S^{(h)})$, the distribution of $S^{(h)}$, 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 $S^{(h)}=\{f(\mathbb{D}_C^{(h)})\}^2$, 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 hence 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 hence remains to be shown that $\mathcal{L}\{f(\mathbb{D}_C^{(h)})\}$ has no atom at a_0. First of all, note that $a_0=0$. Indeed, by Lemma 1.2(e) in Dereich et al. (2003), we have $\mathrm{P}(\|\mathbb{D}_C^{(h)}\|_\infty\le\varepsilon)>0$ 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,

$$\begin{aligned}\mathrm{Var}\{\mathbb{D}_C^{(h)}(\tfrac12,\boldsymbol{u}^*)\}&=\mathrm{Var}\{\tfrac12\mathbb{C}_C^{(h)}(0,1/2,\boldsymbol{u}^*)-\tfrac12\mathbb{C}_C^{(h)}(1/2,1,\boldsymbol{u}^*)\}\\&=\tfrac14\mathrm{Var}\{\mathbb{C}_C^{(h)}(0,1/2,\boldsymbol{u}^*)\}+\tfrac14\mathrm{Var}\{\mathbb{C}_C^{(h)}(1/2,1,\boldsymbol{u}^*)\}=\tfrac14\sigma^2>0.\end{aligned}$$

As consequence, $\mathrm{P}(f(\mathbb{D}_C^{(h)})=0)\le\mathrm{P}(\mathbb{D}_C^{(h)}(\frac12,\boldsymbol{u}^*)=0)=0$, which finally implies that $\mathcal{L}(f(\mathbb{D}_C^{(h)}))$ and therefore $\mathcal{L}(S^{(h)})$ is absolutely continuous. ■

Proof of Proposition 2.4. 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 2.5. The assertions concerning weak convergence are simple consequences of Theorem 1 of Bücher (2015) and of the continuous mapping theorem. Absolute continuity of T can be shown along similar lines as for $S^{(h)}$ in Proposition 2.3. ■

Proof of Proposition A.1. To prove the first claim, one first needs to show that the finite-dimensional distributions of $\big(\mathbb{D}_n^{(h)}, \hat{\mathbb{D}}_n^{(h),[1]}, \dots, \hat{\mathbb{D}}_n^{(h),[M]}, \mathbb{E}_n, \hat{\mathbb{E}}_n^{[1]}, \dots, \hat{\mathbb{E}}_n^{[M]}\big)$ converge weakly to those of $\big(\mathbb{D}_C^{(h)}, \mathbb{D}_C^{(h),[1]}, \dots, \mathbb{D}_C^{(h),[M]}, \mathbb{E}, \mathbb{E}^{[1]}, \dots, \mathbb{E}^{[M]}\big)$. The proof is a more notationally involved version of the proof of Lemma A.1 in Bücher and Kojadinovic (2016a). Joint asymptotic tightness follows from Proposition 2.4 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 (2016a) and the continuous mapping theorem. ■

Proof of Proposition 2.6. As we continue, we adopt the notation $\bar{F}_S^*(x) = \mathrm{P}(\hat{S}_n^{(h),[1]} \geq x \mid \boldsymbol{X}_n)$ and $\bar{F}_T^*(x) = \mathrm{P}(\hat{T}_n^{[1]} \geq x \mid \boldsymbol{X}_n)$, $x \in \mathbb{R}$. Note in passing that the functions $\bar{F}_S^*$ and $\bar{F}_T^*$ are random and that we can rewrite W_n as $W_n = \psi\{\bar{F}_S^*(S_n^{(h)}), \bar{F}_T^*(T_n)\}$. In addition, recall that the distributions of $S^{(h)}$ and T are (absolutely) continuous by Propositions 2.3 and 2.5 and that $\bar{F}_S(x) = \mathrm{P}(S^{(h)} \geq x)$, $x \in \mathbb{R}$, and $\bar{F}_T(x) = \mathrm{P}(T \geq x)$, $x \in \mathbb{R}$. Combining the last statement in Proposition 2.4 with Lemma 2.2 in Bücher and Kojadinovic (2017) and Problem 23.1 in van der Vaart (1998), we obtain that

$$\sup_{x\in\mathbb{R}} |\bar{F}_S^*(x) - \bar{F}_S(x)| \xrightarrow{\mathrm{P}} 0. \tag{B.1}$$

Similarly, by the last claim in Proposition A.1, we obtain that

$$\sup_{x\in\mathbb{R}} |\bar{F}_T^*(x) - \bar{F}_T(x)| \xrightarrow{\mathrm{P}} 0. \tag{B.2}$$

Moreover, Lemma 2.2 in Bücher and Kojadinovic (2017) implies that (B.1) and (B.2) are equivalent to

$$\sup_{x\in\mathbb{R}} \left| \tfrac{1}{M_n} \textstyle\sum_{i=1}^{M_n} \mathbf{1}(S_n^{(h),[i]} \geq x) - \bar{F}_S(x) \right| \xrightarrow{\mathrm{P}} 0, \tag{B.3}$$

and

$$\sup_{x\in\mathbb{R}} \left| \tfrac{1}{M_n} \textstyle\sum_{i=1}^{M_n} \mathbf{1}(T_n^{[i]} \geq x) - \bar{F}_T(x) \right| \xrightarrow{\mathrm{P}} 0, \tag{B.4}$$

respectively.

Now, starting from the last claim in Proposition A.1 and using the continuous mapping theorem, we immediately obtain that, for any $N \in \mathbb{N}$,

$$(\tilde{W}_n, \tilde{W}_n^{[1]}, \dots, \tilde{W}_n^{[N]}) \rightsquigarrow (W, W^{[1]}, \dots, W^{[N]}), \tag{B.5}$$

where $\tilde{W}_n = \psi\{\bar{F}_S(S_n^{(h)}), \bar{F}_T(T_n)\}$ and $\tilde{W}_n^{[i]} = \psi\{\bar{F}_S(\hat{S}_n^{(h),[i]}), \bar{F}_T(\hat{T}_n^{[i]})\}$, $i \in \{1, \dots, N\}$. Some thought reveals that (B.3) and (B.4) combined with (B.5) lead to the fact that (2.23) holds for all $N \in \mathbb{N}$.

From now on, assume that W is continuous. As a straightforward consequence of (B.1) and (B.2), the convergence in (B.5) implies that, for any $N \in \mathbb{N}$,

$$(W_n, W_n^{[1]}, \dots, W_n^{[N]}) \rightsquigarrow (W, W^{[1]}, \dots, W^{[N]}), \tag{B.6}$$

where $W_n^{[i]} = \psi\{\bar{F}_S^*(\hat{S}_n^{(h),[i]}), \bar{F}_T^*(\hat{T}_n^{[i]})\}, i \in \{1, \dots, N\}$. The previous display has the following two consequences: first, by Problem 23.1 in van der Vaart (1998),

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(W_n \leq x) - \mathrm{P}(W \leq x)| \overset{\mathrm{P}}{\to} 0. \tag{B.7}$$

Second, since $W_n^{[1]}, \dots, W_n^{[N]}$ are identically distributed and independent conditionally on the data, by Lemma 2.2 in Bücher and Kojadinovic (2017), we have that (B.6) implies

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(W_n^{[1]} \leq x \,|\, \boldsymbol{X}_n) - \mathrm{P}(W_n \leq x)| \overset{\mathrm{P}}{\to} 0. \tag{B.8}$$

Let us next prove (2.24). In view of (B.8), it suffices to show that

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(W_{n,M_n}^{[1]} \leq x \,|\, \boldsymbol{X}_n) - \mathrm{P}(W_n^{[1]} \leq x \,|\, \boldsymbol{X}_n)| \overset{\mathrm{P}}{\to} 0. \tag{B.9}$$

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

$$|\mathbf{1}(a \leq x) - \mathbf{1}(b \leq x)| \leq \mathbf{1}(|x - a| \leq \varepsilon) + \mathbf{1}(|a - b| > \varepsilon), \tag{B.10}$$

we have that

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

From (B.7) and (B.8), $\sup_{x \in \mathbb{R}} \mathrm{P}(|W_n^{[1]} - x| \leq \varepsilon \,|\, \boldsymbol{X}_n)$ converges in probability to $\sup_{x \in \mathbb{R}} \mathrm{P}(|W - x| \leq \varepsilon)$ which can be made arbitrary small by decreasing ε. From (B.1), (B.3), (B.2) and (B.4), we obtain that $W_n^{[1]} - W_{n,M_n}^{[1]} = o_{\mathrm{P}}(1)$, which implies that

$$\mathrm{P}(|W_n^{[1]} - W_{n,M_n}^{[1]}| > \varepsilon) = \mathrm{E}\{\mathrm{P}(|W_n^{[1]} - W_{n,M_n}^{[1]}| > \varepsilon \,|\, \boldsymbol{X}_n)\} \to 0, \tag{B.11}$$

and thus that $\mathrm{P}(|W_n^{[1]} - W_{n,M_n}^{[1]}| > \varepsilon \,|\, \boldsymbol{X}_n) = o_{\mathrm{P}}(1)$. Hence, (B.9) holds and thus so does (2.24).

Finally, let show that (2.25) holds. Since $W_n^{[1]}, \dots, W_n^{[N]}$ are identically distributed and independent conditionally on the data, by Lemma 2.2 in Bücher and Kojadinovic (2017), we have that (B.8) implies

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}(W_n^{[i]} \leq x) - \mathrm{P}(W_n \leq x) \right| \overset{\mathrm{P}}{\to} 0. \tag{B.12}$$

Whence (2.25) is proved if we show that

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}(W_{n,M_n}^{[i]} \leq x) - \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbf{1}(W_n^{[i]} \leq x) \right| \overset{\mathrm{P}}{\to} 0. \tag{B.13}$$

Using again (B.10), the term on the left of the previous display is smaller than

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

From (B.12) and (B.7), the first term converges in probability to $\sup_{x\in\mathbb{R}} \mathrm{P}(|W-x| \le \varepsilon)$ which can be made arbitrary small by decreasing ε. The second term converges in probability to zero by Markov's inequality: for any $\lambda > 0$,

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

since the $W_n^{[i]} - W_{n,M_n}^{[i]}$ are identically distributed and by (B.11). Therefore, (B.13) holds and, hence, so does (2.25). Note that, from the continuity of $S^{(h)}$, T and W, we could have alternatively proved the analogue statement with '$\le$' replaced by '$<$'. As a consequence, we immediately obtain 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 $\bar{F}_{W_n}(w) = \mathrm{P}(W_n \ge w)$, $w \in \mathbb{R}$. By the analogue to (B.7) with '$\le$' replaced by '$<$', the latter has the same asymptotic distribution as $\bar{F}_W(W_{n,M_n}^{[0]})$, where $\bar{F}_W(w) = \mathrm{P}(W \ge w)$, $w \in \mathbb{R}$. By the weak convergence $W_{n,M_n}^{[0]} \rightsquigarrow W$ following from (2.23) and the continuous mapping theorem, $\bar{F}_W(W_{n,M_n}^{[0]})$ is asymptotically standard uniform. ■

Acknowledgments

Axel Bücher gratefully acknowledges support by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation.

References

Aue, A. and L. Horváth (2013). Structural breaks in time series. *J. Time Series Anal. 34*(1), 1–16.

Auestad, B. and D. Tjøstheim (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika 77*, 669–687.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Chichester: John Wiley and Sons Ltd.

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

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

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

Bücher, A. and I. Kojadinovic (2017). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results.

Bücher, A., I. Kojadinovic, T. Rohmer, and J. Segers (2014). Detecting changes in cross-sectional dependence in multivariate time series. *Journal of Multivariate Analysis 132*, 111–128.

Bücher, A. and S. Volgushev (2013). Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis 119*, 61–70.

Cardinali, A. and G. Nason (2013). Costationarity of Locally Stationary Time Series Using `costat`. *Journal of Statistical Software 55*(1), 1–22.

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

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

Chernozhukov, V., D. Chetverikov, and K. Kato (2013, 12). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist. 41*(6), 2786–2819.

Constantine, W. and D. Percival (2016). *fractal: Fractal time series modeling and analysis.* R package version 2.0-1.

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

Davydov, Y. A. and M. A. Lifshits (1984). The fibering method in some probability problems. In *Probability theory. Mathematical statistics. Theoretical cybernetics, Vol. 22*, Itogi Nauki i Tekhniki, pp. 61–157, 204. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow.

Dedecker, J., F. Merlevède, and E. Rio (2014). Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electron. J. Probab. 19*, no. 9, 56.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. 5th Ser. 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., F. Fehringer, A. Matoussi, and M. Scheutzow (2003). On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *J. Theoret. Probab. 16*(1), 249–265.

Dette, H., P. Preuss, and M. Vetter (2011). A measure of stationarity in locally stationary processes with applications to testing. *J. Amer. Statist. Assoc. 106*, 1113–1124.

Dette, H., W. Wu, and Z. Zhou (2015). Change point analysis of second order characteristics in non-stationary time series.

Dwivedi, Y. and S. Subba Rao (2011). A test for second-order stationarity of a time series based on the discrete fourier transform. *J. Time Series Anal. 32*, 68–91.

Escanciano, J. C. and I. N. Lobato (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. and B. Rémillard (2004). Tests of independence and randomness based on the empirical copula process. *Test 13*(2), 335–369.

Gombay, E. and L. Horváth (1999). Change-points and bootstrap. *Environmetrics 10*(6).

Grégoire, V., C. Genest, and M. Gendron (2008). Using copulas to model price dependence in energy markets. *Energy Risk 5*(5), 58–64.

Hofert, M., I. Kojadinovic, M. Mächler, and J. Yan (2017). *copula: Multivariate dependence with copulas.* R package version 0.999-17.

Holmes, M., I. Kojadinovic, and J.-F. Quessy (2013). Nonparametric tests for change-point detection à la Gombay and Horváth. *Journal of Multivariate Analysis 115*, 16–32.

Hušková, M. and A. Slabý (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., S. Wang, and H. Wang (2015). A new non-parametric stationarity test of time series in the time domain. *J. R. Stat. Soc. Ser. B 77*, 893–922.

Jondeau, E., S.-H. Poon, and M. Rockinger (2007). *Financial modeling under non-Gaussian distributions.* London: Springer.

Kirch, C. and B. Muhsal (2016). A mosum procedure for the estimation of multiple random change points. *Bernoulli, to appear*.

Kojadinovic, I. (2017). *npcp: Some Nonparametric Tests for Change-Point Detection in Possibly Multivariate Observations.* R package version 0.1-9.

Kosorok, M. (2008). *Introduction to empirical processes and semiparametric inference.* New York: Springer.

Künsch, H. (1989). The jacknife and the bootstrap for general stationary observations. *The Annals of Statistics 17*(3), 1217–1241.

Lee, S., J. Ha, and O. Na (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics 30*, 781–796.

McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative risk management.* New Jersey: Princeton University Press.

Nason, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Stat. Soc. Ser. B 75*, 879–904.

Nason, G. (2016). *locits: Tests of stationarity and localized autocovariance.* R package version 1.7.3.

Nason, G. and A. Cardinali (2013). *costat: Time series costationarity determination.* R package version 2.3.

Nelsen, R. (2006). *An introduction to copulas.* New-York: Springer. Second edition.

Nualart, D. (2006). *The Malliavin calculus and related topics* (Second ed.). Probability and its Applications (New York). Springer-Verlag, 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. *J. Amer. Statist. 105*, 839–851.

Paparoditis, E. and D. Politis (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., D. Politis, and H. White (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. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews 23*(1), 53–70.

Preuss, P., M. Vetter, and H. Dette (2013). A test for stationarity based on empirical processes. *Bernoulli 19*(5B), 2715–2749.

Priestley, M. and T. Subba Rao (1969). A test for non-stationarity of time-series. *J. R. Stat. Soc. Ser. B 31*, 140–149.

Puchstein, R. and P. Preuss (2016). Testing for stationarity in multivariate locally stationary processes. *J. Time Series Anal. 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 multivariate distributions with given marginals. *Ann. Inst. Statist. Math. 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., E. Suchman, L. DeVinney, S. Star, and R. J. Williams (1949). *The American Soldier, Vol.1: Adjustment during Army Life.* Princeton: Princeton University Press.

van der Vaart, A. (1998). *Asymptotic statistics.* Cambridge University Press.

van der Vaart, A. and J. Wellner (2000). *Weak convergence and empirical processes.* New York: Springer. Second edition.

von Sachs, R. and M. Neumann (2000). A wavelet-based test for stationarity. *J. Time Series Anal. 21*, 597–613.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc. 108*, 726–740.

Table 2: Percentage of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n \in \{128, 256, 512\}$ generated from Models N1–N5, first with $N(0,1)$ innovations and then with standardized t_4 innovations. The meaning of the abbreviations d, m, v, c, dc, dcp, a, va and mva is given in Section 5.

						$h=2$ or lag 1					$h=3$ or lag 2					$h=4$ or lag 3					$h=8$	
Model	Innov.	n	d	m	v	c	dc	a	va	mva	c	dc	dcp	va	mva	c	dc	dcp	va	mva	c	dc
N1	N(0,1)	128	4.0	4.7	3.8	3.0	3.9	3.0	3.5	3.6	4.6	4.1	4.2	3.9	4.1	3.2	3.4	5.7	4.1	4.5	0.0	0.7
		256	4.9	5.3	4.1	4.4	4.1	4.5	4.8	4.6	5.5	5.2	5.9	4.4	5.1	4.2	4.7	5.9	4.4	4.9	0.0	0.5
		512	5.8	4.7	4.6	4.6	5.0	4.3	5.5	4.4	6.2	6.6	5.0	5.4	4.3	6.6	6.3	5.2	4.9	5.0	0.2	2.3
	Stand. t_4	128	4.5	3.4	1.6	4.2	5.0	5.2	3.7	3.3	4.9	5.3	5.7	4.7	3.0	3.4	4.0	5.9	3.7	2.4	0.0	0.7
		256	5.1	3.8	1.4	4.2	5.9	4.2	3.6	4.0	5.9	5.5	5.4	3.0	2.6	5.3	5.4	5.4	2.8	3.0	0.0	0.4
		512	4.7	4.5	2.6	5.8	5.7	4.9	3.8	4.0	8.6	6.7	4.5	4.0	3.1	7.1	6.0	5.9	3.2	3.1	0.3	1.2
N2	N(0,1)	128	2.5	2.5	1.3	1.6	3.3	1.3	4.9	3.2	1.6	2.6	5.1	7.5	3.8	1.5	2.0	6.4	8.7	4.1	0.2	0.0
		256	5.1	4.2	1.8	2.3	3.5	1.3	5.7	3.0	2.2	3.9	5.5	8.0	4.4	2.1	3.6	6.3	9.1	4.4	0.5	1.8
		512	5.1	5.0	1.5	1.5	2.2	1.7	4.4	4.2	1.9	3.1	3.8	6.9	5.2	1.4	3.1	4.8	7.5	5.3	0.7	2.5
	Stand. t_4	128	2.4	2.0	1.1	2.1	3.4	1.0	4.4	3.0	1.3	3.0	5.2	6.7	3.3	1.6	2.2	5.7	7.5	3.9	0.0	0.4
		256	5.6	4.8	1.2	1.0	1.3	1.1	4.5	2.8	0.8	1.9	4.4	7.3	4.0	0.5	2.0	4.9	8.3	4.4	0.9	1.3
		512	6.1	5.9	1.9	0.6	1.9	1.4	4.5	3.5	1.0	2.7	4.1	6.4	4.5	1.1	2.7	5.0	7.6	4.7	0.9	2.8
N3	N(0,1)	128	0.5	0.0	0.5	2.2	3.2	0.5	2.8	0.7	0.7	0.9	6.5	3.3	0.7	0.0	0.0	6.3	4.1	1.0	0.0	0.0
		256	1.4	0.0	0.5	2.4	4.1	0.6	2.3	0.7	0.1	1.0	6.7	3.6	1.0	0.0	0.0	6.6	4.0	1.2	0.0	0.0
		512	2.7	0.3	1.4	2.2	4.8	1.1	4.2	1.6	0.5	2.0	7.7	5.9	1.9	0.2	0.5	7.9	6.9	2.1	0.0	0.0
	Stand. t_4	128	0.4	0.1	0.4	2.0	3.7	0.1	1.8	0.4	0.1	0.3	5.0	2.5	0.3	0.0	0.0	5.4	2.6	0.2	0.0	0.0
		256	2.2	0.1	0.8	2.7	5.3	0.4	2.8	0.6	0.4	2.0	7.9	4.4	0.8	0.0	0.3	8.5	5.4	0.8	0.0	0.0
		512	2.8	0.7	1.0	2.9	6.3	0.6	3.4	1.8	0.3	2.0	7.7	5.2	2.0	0.1	0.7	8.7	5.2	2.2	0.0	0.2
N4	N(0,1)	128	4.8	5.3	3.9	2.5	4.2	2.5	7.3	5.5	3.3	4.5	7.3	7.0	4.9	3.7	4.6	6.6	6.9	4.7	0.1	1.1
		256	5.7	7.3	3.5	2.5	4.6	3.7	7.9	7.7	4.0	5.3	7.0	8.2	6.6	4.9	5.7	6.5	6.8	6.4	0.9	1.9
		512	5.6	6.6	4.6	4.9	5.4	4.7	7.1	7.9	6.2	5.8	6.9	7.4	7.2	6.6	5.6	7.3	7.4	6.3	1.9	3.3
	Stand. t_4	128	3.4	3.9	2.7	3.8	4.6	1.2	4.5	2.8	3.5	5.0	6.4	6.2	3.7	3.5	4.7	6.3	6.4	3.4	0.2	1.4
		256	4.4	3.6	1.5	2.4	3.7	1.5	4.3	2.3	3.8	5.0	5.9	5.4	3.1	4.9	5.5	6.4	4.8	2.5	0.3	0.8
		512	5.2	4.7	2.7	4.4	5.2	1.5	5.8	4.5	5.8	5.5	6.4	5.9	4.7	6.6	6.0	6.8	5.8	5.0	2.3	4.5
N5	N(0,1)	128	1.3	0.0	1.1	3.2	2.0	1.3	3.4	0.6	2.2	1.8	2.1	3.1	0.5	0.3	0.2	2.2	2.7	0.2	0.0	0.0
		256	2.0	0.0	1.3	3.5	3.4	1.3	3.5	0.5	2.6	2.7	3.6	3.6	0.1	0.8	1.3	3.2	3.2	0.0	0.0	0.2
		512	3.2	0.1	2.0	4.7	4.4	3.1	5.9	1.6	3.6	3.7	5.0	4.9	1.3	1.7	2.6	4.6	4.9	0.9	0.0	0.2
	Stand. t_4	128	1.4	0.0	1.1	3.8	3.0	1.6	4.3	0.4	2.5	1.7	3.3	3.9	0.2	0.5	0.6	3.3	3.2	0.1	0.0	0.0
		256	2.4	0.0	0.7	3.9	2.9	1.5	3.4	0.3	2.6	2.1	4.2	3.2	0.1	0.8	0.8	2.9	2.6	0.0	0.0	0.2
		512	3.9	0.0	1.1	4.0	4.0	1.3	4.0	0.6	3.6	4.3	5.1	3.2	0.3	1.5	1.8	5.1	2.9	0.1	0.0	0.7

Table 3: Continued from Table 2. Percentage of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n \in \{128, 256, 512\}$ generated from Models N6–N10, first with $N(0,1)$ innovations and then with standardized t_4 innovations. The meaning of the abbreviations d, m, v, c, dc, dcp, a, va and mva is given in Section 5.

						$h=2$ or lag 1					$h=3$ or lag 2					$h=4$ or lag 3					$h=8$	
Model	Innov.	n	d	m	v	c	dc	a	va	mva	c	dc	dcp	va	mva	c	dc	dcp	va	mva	c	dc
N6	N(0,1)	128	2.5	0.1	1.1	3.4	3.7	1.2	4.1	1.0	0.2	0.9	5.0	5.8	1.2	0.0	0.0	5.0	5.5	1.1	0.0	0.0
		256	4.1	1.1	2.9	2.7	5.5	2.4	7.9	3.1	1.3	3.3	7.2	8.6	3.9	0.0	0.5	6.8	7.8	3.2	0.0	0.2
		512	5.5	1.6	3.8	4.2	5.8	4.3	10.3	6.5	1.5	4.0	7.1	12.1	7.1	0.2	1.1	8.2	11.8	5.8	0.0	0.8
	Stand. t_4	128	1.8	0.1	1.1	2.6	3.0	1.2	4.0	0.9	0.1	0.9	4.5	4.4	1.2	0.0	0.2	4.9	3.8	0.8	0.0	0.0
		256	4.1	0.7	2.2	2.7	4.3	1.5	5.1	2.4	1.1	2.3	5.9	7.1	2.3	0.2	0.9	6.2	6.2	1.7	0.0	0.5
		512	4.6	1.5	2.5	4.7	5.5	2.2	6.9	3.2	1.1	2.5	6.9	8.4	3.4	0.0	0.6	7.1	7.5	3.1	0.0	0.5
N7	N(0,1)	128	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.4	0.4	0.8	2.1	0.0	3.5	3.4	5.1	2.8	0.0	0.1	0.1
		256	1.8	0.0	0.1	0.0	0.0	0.2	3.8	0.1	0.0	0.7	0.7	2.2	0.1	0.8	1.6	2.2	4.3	0.1	0.1	0.5
		512	6.8	0.0	0.9	0.0	0.5	0.8	4.9	0.9	0.0	2.1	3.2	2.3	0.1	1.0	3.7	5.2	5.3	0.4	0.2	1.0
	Stand. t_4	128	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.6	0.3	0.9	2.2	0.0	2.8	2.4	4.0	2.7	0.0	0.0	0.0
		256	1.6	0.0	0.3	0.0	0.1	0.0	3.6	0.2	0.0	0.3	0.6	1.4	0.0	0.9	1.4	3.3	4.2	0.1	0.2	0.2
		512	6.5	0.0	0.6	0.0	1.0	0.9	3.2	0.7	0.0	2.1	3.1	2.5	0.4	0.5	3.2	5.2	3.5	0.7	0.3	0.8
N8	N(0,1)	128	5.9	5.0	23.5	4.4	6.0	4.3	17.7	11.5	4.9	5.5	6.3	22.4	11.6	3.0	4.2	7.0	23.4	12.3	0.0	0.7
		256	6.3	5.7	31.9	5.8	6.6	4.0	26.2	19.1	5.6	6.4	6.7	30.8	24.1	4.3	5.6	6.9	31.9	23.2	0.0	1.2
		512	7.1	5.5	37.2	5.7	6.4	4.1	30.6	25.4	7.9	8.8	7.7	33.0	27.7	8.0	8.5	7.4	35.5	28.5	0.5	2.4
	Stand. t_4	128	6.0	3.6	13.0	4.0	4.4	4.3	11.2	8.0	4.6	5.8	4.9	16.5	7.5	3.2	4.9	6.0	15.6	7.9	0.0	0.7
		256	6.8	4.3	19.5	3.4	5.7	2.8	14.4	10.4	6.2	7.1	6.6	20.2	12.3	5.2	6.1	7.7	21.5	12.3	0.0	1.5
		512	8.2	4.4	23.6	4.1	6.3	3.1	19.6	15.2	6.0	8.0	7.3	21.3	15.9	6.2	7.9	8.3	24.4	16.0	0.2	2.2
N9	N(0,1)	128	5.0	4.6	1.8	2.6	4.0	0.9	5.3	2.5	2.7	3.8	5.2	6.0	2.4	2.2	3.4	6.5	4.7	2.2	0.0	0.5
		256	5.6	4.8	1.9	2.3	4.4	2.4	4.9	3.6	2.5	4.2	6.2	6.4	3.6	2.4	4.2	6.4	6.5	3.3	0.1	1.3
		512	4.8	4.1	1.4	3.0	4.0	1.0	4.8	3.0	2.9	4.2	5.3	6.3	3.5	2.6	4.2	5.1	6.7	3.5	0.7	2.3
	Stand. t_4	128	3.0	2.1	1.7	2.4	3.5	1.0	4.7	1.9	2.4	2.8	4.1	5.4	1.9	1.4	2.5	4.5	5.5	2.0	0.0	0.3
		256	5.5	4.6	1.4	2.5	2.9	1.1	3.6	2.6	2.1	3.8	4.9	5.4	2.9	1.3	3.4	4.9	5.1	2.5	0.1	0.7
		512	4.6	4.2	1.8	2.9	4.3	1.6	5.1	3.9	3.3	5.2	5.8	7.0	4.3	3.2	4.8	6.8	7.3	4.1	0.5	2.8
N10	N(0,1)	128	5.9	5.9	3.9	3.7	5.2	4.0	5.2	4.5	3.8	5.8	6.5	5.7	4.7	3.1	5.9	7.1	5.1	4.4	0.2	0.8
		256	6.6	6.8	3.3	3.5	5.1	3.0	5.1	4.7	3.3	5.9	6.8	6.2	5.2	3.9	5.3	7.5	5.3	5.2	0.6	1.6
		512	7.4	6.7	3.8	4.5	6.6	4.0	5.4	6.8	5.0	7.2	7.8	6.1	6.7	5.4	6.6	8.6	6.1	6.3	3.1	3.6
	Stand. t_4	128	7.0	7.8	1.6	4.0	6.1	4.8	4.6	6.1	4.4	5.9	7.7	5.5	5.4	4.0	5.9	9.4	4.1	5.5	0.1	1.7
		256	6.9	5.3	2.0	3.6	6.1	3.9	3.8	4.1	4.0	5.9	7.9	2.7	4.1	4.8	6.6	8.2	3.0	4.5	0.4	2.3
		512	6.7	5.8	2.2	3.4	5.6	3.6	3.3	4.0	5.0	6.1	6.4	3.7	4.3	5.3	6.1	7.2	3.0	4.4	4.2	4.0

Table 4: Percentage of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n \in \{128, 256, 512\}$ generated from Models A1–A8. The meaning of the abbreviations d, m, v, c, dc, dcp, a, va and mva is given in Section 5.

					$h = 2$ or lag 1					$h = 3$ or lag 2					$h = 4$ or lag 3					$h = 8$	
Model	n	d	m	v	c	dc	a	va	mva	c	dc	dcp	va	mva	c	dc	dcp	va	mva	c	dc
A1	128	3.3	2.5	4.3	5.0	4.3	5.5	6.5	5.1	6.7	5.2	5.5	6.5	4.0	3.2	3.8	6.2	6.3	4.6	0.2	0.9
	256	5.5	5.0	5.2	11.4	10.5	12.9	12.9	10.4	13.2	11.0	9.1	11.8	8.8	6.9	6.7	9.1	10.9	8.5	0.2	1.2
	512	4.2	4.4	6.4	42.5	27.5	37.7	30.9	23.1	42.5	26.6	16.0	19.9	14.9	25.2	15.6	10.6	15.3	11.5	1.4	3.5
A2	128	4.9	5.2	2.1	9.7	9.3	13.2	10.1	9.5	10.1	8.2	8.7	9.2	8.2	4.7	5.0	8.3	7.1	6.4	0.0	0.8
	256	6.1	6.3	3.9	36.1	26.2	34.9	26.2	22.2	31.9	23.5	15.0	16.9	14.6	17.2	14.3	12.1	14.4	11.4	0.9	2.1
	512	5.6	5.5	4.1	76.1	58.2	61.6	40.6	34.1	72.4	51.0	28.1	21.6	18.1	54.8	37.6	18.1	17.4	13.0	8.7	7.1
A3	128	6.2	6.3	4.3	13.1	12.9	18.8	16.9	15.1	11.8	11.9	10.4	13.5	11.4	6.0	7.2	10.1	11.7	10.1	0.1	1.1
	256	5.8	5.4	4.9	42.1	30.6	40.7	28.3	23.6	36.1	25.2	17.6	18.7	14.3	19.5	16.1	13.8	15.1	11.0	1.4	2.3
	512	5.7	6.0	4.6	92.4	77.0	83.8	63.6	51.0	89.3	71.4	36.9	30.7	23.5	75.2	54.4	25.8	23.4	16.8	15.9	11.5
A4	128	2.0	1.9	1.8	3.3	2.5	2.4	4.9	2.9	3.0	2.5	4.1	5.7	3.0	1.4	1.2	3.8	6.1	2.9	0.0	0.0
	256	4.0	2.0	1.7	3.4	3.4	2.2	5.1	2.9	3.3	3.2	4.3	6.1	3.0	1.6	1.9	4.8	5.4	2.5	0.0	0.2
	512	3.4	2.1	2.3	3.9	5.1	2.5	5.8	3.7	4.0	4.9	6.0	6.5	3.7	2.8	4.0	5.5	6.8	3.4	0.0	0.7
A5	128	4.9	5.6	3.2	64.0	49.9	45.6	44.2	37.1	53.4	44.5	28.3	31.8	24.7	38.5	29.9	19.8	26.7	20.3	1.2	2.5
	256	5.6	5.4	4.0	84.5	75.5	60.4	54.4	43.6	77.6	68.2	46.2	40.9	29.9	63.6	52.1	35.8	36.6	25.6	12.1	9.5
	512	4.9	4.4	8.4	96.1	91.5	76.5	72.4	63.2	93.0	86.8	75.0	64.6	46.5	84.4	74.9	67.2	62.9	42.1	38.8	31.2
A6	128	1.2	0.0	0.1	3.8	3.6	0.1	0.4	0.0	2.3	2.1	3.2	0.3	0.0	0.4	0.6	3.4	0.5	0.0	0.0	0.0
	256	7.1	0.0	0.3	4.6	8.2	0.3	1.4	0.0	3.9	6.9	10.4	1.6	0.0	0.7	2.7	10.5	1.9	0.0	0.0	0.4
	512	45.6	0.0	1.2	5.1	28.0	1.4	5.3	0.7	4.2	27.9	39.2	4.6	0.3	2.4	18.3	45.0	5.3	0.3	0.0	5.8
A7	128	0.2	0.0	2.5	3.9	0.9	2.0	3.4	0.4	6.9	1.8	0.5	4.2	0.7	4.7	1.3	0.7	3.5	0.5	0.0	0.0
	256	0.8	0.0	1.4	6.5	1.8	2.2	2.3	0.2	6.4	1.8	1.8	2.8	0.3	4.9	1.8	2.2	2.7	0.2	0.0	0.1
	512	2.8	0.0	2.0	19.9	10.3	1.9	1.9	0.1	18.1	9.3	8.8	4.3	0.2	12.7	6.5	7.1	3.6	0.1	0.0	0.2
A8	128	0.0	0.0	12.2	12.5	4.0	7.5	15.8	7.2	20.0	7.3	3.6	18.0	4.3	23.8	9.6	5.4	24.3	5.9	7.8	2.0
	256	0.0	0.0	12.1	30.4	11.4	11.3	19.9	6.4	39.3	21.7	16.5	25.6	6.2	45.3	27.8	21.3	34.9	10.4	42.4	20.7
	512	0.4	0.0	16.4	37.6	26.6	16.8	24.8	17.1	54.2	34.8	38.3	26.3	17.1	73.6	46.7	52.7	30.6	22.5	87.3	57.2

Table 5: Percentage of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n \in \{128, 256, 512\}$ generated from Model A9 with $\beta \in \{-0.8, -0.4, 0, 0.4, 0.8\}$. The meaning of the abbreviations d, m, v, c, dc, dcp, a, va and mva is given in Section 5.

						$h = 2$ or lag 1					$h = 3$ or lag 2					$h = 4$ or lag 3				
Innov.	n	β	d	m	v	c	dc	a	va	mva	c	dc	dcp	va	mva	c	dc	dcp	va	mva
N(0,1)	128	-0.8	10.0	1.6	14.4	61.2	59.5	42.1	52.5	37.7	49.8	52.9	58.8	56.2	36.4	20.6	25.8	49.6	54.3	33.0
		-0.4	4.6	4.8	4.6	32.0	22.6	34.2	28.2	20.2	26.3	18.7	14.3	21.5	13.3	14.9	10.4	11.5	15.9	9.8
		0.0	3.9	3.1	3.8	3.6	4.2	3.9	4.2	3.8	5.3	5.1	5.2	4.6	2.7	3.8	3.7	5.9	4.3	2.7
		0.4	8.6	8.3	3.9	28.0	26.6	31.2	24.1	20.8	27.9	25.2	19.0	15.5	12.2	17.8	16.7	16.4	12.5	9.8
		0.8	6.7	6.1	12.7	52.8	45.2	39.5	47.5	32.4	49.2	43.2	41.2	50.9	32.1	39.6	34.4	33.4	45.9	27.8
	256	-0.8	40.9	2.5	46.1	95.5	97.0	74.2	81.8	69.5	93.5	95.0	96.6	83.1	70.7	69.4	84.8	95.2	82.8	67.6
		-0.4	5.7	4.6	10.2	66.4	52.8	69.0	60.3	48.6	57.3	44.7	34.9	45.0	33.0	35.4	28.2	23.3	32.1	22.3
		0.0	3.7	4.0	4.6	4.4	3.3	4.7	4.1	4.1	6.2	5.1	4.6	4.9	4.8	5.3	4.2	4.8	5.1	4.8
		0.4	6.8	6.6	6.6	63.6	54.9	69.8	57.3	48.4	60.8	51.7	37.6	39.7	28.7	50.6	42.3	25.1	27.3	19.5
		0.8	10.0	6.9	37.6	89.4	85.0	74.7	81.2	68.2	86.9	82.2	80.8	84.0	67.0	80.5	76.5	72.1	82.0	63.3
	512	-0.8	87.7	2.0	85.9	100.0	100.0	95.2	97.7	93.9	100.0	100.0	100.0	98.3	94.7	99.2	99.9	99.9	98.4	93.7
		-0.4	6.8	5.6	17.3	95.7	89.8	98.1	93.8	88.6	89.6	83.9	75.8	81.4	67.2	75.2	62.9	48.8	63.2	48.4
		0.0	4.2	4.5	4.2	5.2	5.2	5.0	4.1	4.0	6.5	5.6	4.9	4.2	4.7	6.3	4.9	5.1	4.6	4.9
		0.4	7.0	5.5	12.8	95.0	89.7	96.2	90.5	86.4	91.9	86.4	74.9	77.8	63.4	86.8	78.2	50.5	60.0	48.0
		0.8	11.2	6.8	84.2	99.7	99.7	96.1	98.5	94.3	99.6	99.5	99.5	99.1	94.4	99.0	98.3	98.8	98.6	93.7
St. t_4	128	-0.8	15.8	1.0	7.7	61.5	67.1	34.4	37.6	23.5	51.5	59.9	67.7	42.5	23.5	21.8	35.1	60.3	41.0	22.1
		-0.4	4.4	2.5	2.6	38.3	27.8	25.9	21.0	11.2	33.5	23.5	16.4	13.2	6.1	15.9	11.7	12.0	10.6	4.3
		0.0	5.1	3.6	1.2	3.2	4.6	4.4	4.1	2.8	5.8	3.9	5.3	3.9	1.9	3.9	3.2	6.1	3.5	1.7
		0.4	9.4	7.0	2.1	37.3	32.0	26.1	19.8	14.2	32.0	29.1	23.8	12.5	8.9	22.3	21.2	21.0	10.6	6.7
		0.8	9.4	7.4	7.9	56.3	52.9	33.5	37.2	27.3	50.2	48.0	47.2	39.4	25.2	39.2	38.9	39.7	36.8	21.1
	256	-0.8	56.3	2.2	26.6	96.1	98.5	65.3	69.8	54.7	95.8	98.6	97.7	73.4	56.2	73.3	92.2	96.9	72.1	53.0
		-0.4	5.4	3.0	3.9	74.5	61.4	58.9	44.6	31.9	65.2	54.4	43.3	30.3	18.1	41.5	32.4	28.0	21.5	12.3
		0.0	4.7	4.8	1.7	4.8	4.1	4.8	3.5	3.2	6.9	6.4	5.1	2.2	2.5	6.6	6.2	5.3	2.3	2.4
		0.4	6.4	5.6	3.2	71.5	62.9	58.1	45.3	35.6	67.7	57.8	44.4	27.8	19.4	56.5	47.1	30.6	18.1	13.2
		0.8	11.2	6.1	20.6	92.4	87.6	62.1	65.6	51.2	89.2	86.1	84.2	68.7	51.5	83.9	81.0	77.0	66.0	47.5
	512	-0.8	96.1	2.8	58.4	100.0	100.0	88.4	92.3	83.9	100.0	100.0	100.0	93.9	84.6	99.9	100.0	100.0	93.8	84.0
		-0.4	8.8	4.7	5.1	98.6	96.2	89.3	81.5	69.7	95.3	91.3	85.2	65.7	46.4	84.6	74.8	60.0	46.3	27.3
		0.0	4.9	3.3	1.6	5.1	5.6	4.3	3.9	5.1	6.7	6.2	5.4	3.9	3.4	7.2	6.2	6.0	2.4	2.3
		0.4	7.0	5.4	4.9	98.1	95.8	87.8	79.3	69.7	97.2	92.8	81.4	62.6	46.1	93.1	85.4	59.5	40.8	28.1
		0.8	14.9	5.7	57.1	99.9	100.0	87.5	89.7	82.5	99.7	99.6	99.9	91.7	83.1	99.7	99.4	99.5	91.9	82.2

Table 6: Percentage of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n \in \{128, 256, 512\}$ generated from Model A10 with $\beta \in \{-0.8, -0.4, 0, 0.4, 0.8\}$. The meaning of the abbreviations d, m, v, c, dc, dcp, a, va and mva is given in Section 5.

						$h = 2$ or lag 1					$h = 3$ or lag 2					$h = 4$ or lag 3				
Innov.	n	β	d	m	v	c	dc	a	va	mva	c	dc	dcp	va	mva	c	dc	dcp	va	mva
N(0,1)	128	-0.8	2.6	0.4	11.6	1.5	3.7	2.8	10.2	4.4	16.1	21.9	26.7	35.4	15.3	11.3	16.2	20.0	29.9	11.0
		-0.4	2.4	2.5	7.0	3.7	2.6	2.3	4.2	2.6	9.3	5.6	9.3	19.4	9.8	8.7	4.9	6.2	15.5	7.2
		0.0	3.9	3.1	3.8	3.6	4.2	3.9	4.2	3.8	5.3	5.1	5.2	4.6	2.7	3.8	3.7	5.9	4.3	2.7
		0.4	8.8	11.3	4.3	5.2	9.1	6.2	6.2	7.8	9.0	11.3	21.0	19.0	15.8	7.4	10.1	17.3	14.8	12.7
		0.8	7.3	8.6	11.3	1.5	5.6	4.1	14.1	13.4	2.5	5.6	25.2	38.7	27.4	2.8	5.0	22.7	36.7	23.9
	256	-0.8	28.8	0.8	33.9	1.2	13.0	1.3	18.5	10.0	67.5	80.8	74.7	63.8	38.0	52.5	73.3	61.4	56.6	30.9
		-0.4	4.0	3.2	8.8	3.4	4.2	3.2	6.4	4.3	21.6	15.4	24.9	39.0	25.3	20.4	13.9	15.8	27.5	18.1
		0.0	3.7	4.0	4.6	4.4	3.3	4.7	4.1	4.1	6.2	5.1	4.6	4.9	4.8	5.3	4.2	4.8	5.1	4.8
		0.4	9.1	9.2	8.9	4.7	7.9	6.0	9.2	10.8	11.6	15.6	33.2	41.5	35.8	14.5	17.1	25.6	30.2	26.0
		0.8	10.4	7.6	36.2	1.1	6.1	3.2	27.2	21.2	4.4	8.4	41.8	71.0	53.3	5.8	11.1	31.1	66.1	45.2
	512	-0.8	81.3	1.5	78.2	2.9	56.0	1.9	53.0	35.4	99.5	99.8	98.7	90.7	75.7	97.4	99.5	97.0	88.1	69.3
		-0.4	4.1	3.6	17.4	3.7	3.0	4.4	10.9	7.7	47.8	35.3	58.9	74.0	51.7	49.0	35.3	32.7	53.4	35.1
		0.0	4.2	4.5	4.2	5.2	5.2	5.0	4.1	4.0	6.5	5.6	4.9	4.2	4.7	6.3	4.9	5.1	4.6	4.9
		0.4	8.0	7.1	15.8	4.2	7.8	5.3	11.7	11.8	23.7	20.7	62.5	72.2	57.6	32.4	25.9	36.0	51.9	40.0
		0.8	17.1	8.1	77.7	1.4	8.3	3.8	60.4	46.3	16.2	27.7	83.2	93.8	84.2	27.8	40.3	58.5	91.9	78.3
St. t_4	128	-0.8	5.7	0.4	6.9	1.4	5.1	4.4	7.6	4.5	17.0	29.5	36.0	26.6	10.2	11.5	21.6	28.1	21.7	6.5
		-0.4	2.0	1.9	2.5	2.8	1.8	4.2	4.3	3.2	13.4	6.7	12.2	14.4	6.9	10.1	6.6	8.1	11.1	4.5
		0.0	5.1	3.6	1.2	3.2	4.6	4.4	4.1	2.8	5.8	3.9	5.3	3.9	1.9	3.9	3.2	6.1	3.5	1.7
		0.4	11.8	10.3	2.8	5.5	9.8	7.4	7.7	8.4	8.0	11.6	23.9	15.9	14.9	8.2	10.9	21.5	11.6	11.5
		0.8	9.9	7.9	8.5	1.6	7.1	8.4	16.0	14.5	1.9	6.4	30.3	35.1	22.7	1.6	5.4	26.1	33.4	22.4
	256	-0.8	44.7	0.2	19.7	0.6	20.1	3.2	12.5	5.9	69.4	88.5	83.0	46.8	24.9	55.7	82.3	74.3	40.4	18.4
		-0.4	3.0	1.9	4.3	3.9	3.0	2.9	4.5	2.8	23.3	16.0	28.8	27.8	15.1	22.9	14.9	16.3	17.8	7.9
		0.0	4.7	4.8	1.7	4.8	4.1	4.8	3.5	3.2	6.9	6.4	5.1	2.2	2.5	6.6	6.2	5.3	2.3	2.4
		0.4	10.0	9.1	2.9	4.4	9.1	5.6	6.5	7.8	11.0	15.3	41.3	28.0	21.0	14.8	17.6	28.9	18.2	15.7
		0.8	13.7	7.9	20.3	1.3	6.7	5.1	18.5	14.3	3.0	8.3	50.1	53.4	39.2	4.7	11.7	35.8	49.0	32.5
	512	-0.8	92.3	2.1	52.5	2.9	70.6	2.1	29.2	18.8	99.8	100.0	99.8	77.7	55.1	98.3	99.7	98.6	72.0	46.4
		-0.4	7.5	2.6	8.1	4.3	6.0	3.3	5.8	4.9	60.1	49.9	69.8	55.7	32.6	57.6	47.5	41.9	33.5	18.9
		0.0	4.9	3.3	1.6	5.1	5.6	4.3	3.9	5.1	6.7	6.2	5.4	3.9	3.4	7.2	6.2	6.0	2.4	2.3
		0.4	10.5	8.3	6.5	6.0	8.2	5.4	6.5	6.6	27.1	28.2	74.4	53.6	39.7	36.0	33.6	49.4	33.0	25.7
		0.8	19.1	6.7	51.8	1.1	9.7	4.3	37.7	27.8	14.2	28.3	86.5	80.7	64.6	27.1	42.7	64.2	75.0	57.7

Table 7: Percentage of rejection of the null hypothesis of stationarity computed from 1000 samples of size $n \in \{128, 256, 512\}$ generated from Models A11 and A12 with $\beta \in \{0, 0.4, 0.8\}$. The meaning of the abbreviations d, c, dc, dcp is given in Section 5. The columns c2 and c3 report the results for the bivariate analogues of the test based on $S_n^{(2)}$ defined by (2.3) for lags 2 and 3 (these tests arise in the combined test dcp).

				$h = 2$ or lag 1		$h = 3$ or lag 2				$h = 4$ or lag 3			
Model	n	β	d	c	dc	c	dc	c2	dcp	c	dc	c3	dcp
A11	128	0.0	4.3	4.9	4.5	4.8	3.8	5.2	5.1	4.1	3.2	6.4	5.1
		0.4	8.3	69.3	60.7	60.6	53.3	8.4	41.8	48.6	40.4	7.1	30.3
		0.8	22.1	91.9	90.4	86.4	84.6	80.0	91.2	80.8	79.5	55.5	88.6
	256	0.0	4.3	3.8	4.3	6.3	5.9	4.8	5.0	5.9	4.9	5.6	5.5
		0.4	6.4	99.5	96.1	96.9	91.2	20.0	78.4	92.3	84.5	6.2	54.7
		0.8	25.3	99.3	98.5	96.4	94.9	95.8	98.6	92.4	90.4	85.9	96.8
	512	0.0	4.6	6.6	5.9	7.0	6.5	4.2	5.2	5.6	6.2	5.0	5.0
		0.4	6.4	100.0	100.0	99.9	100.0	56.7	99.2	99.8	99.6	10.7	83.6
		0.8	22.1	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.7	99.7	100.0
A12	128	0.0	5.6	3.7	4.7	4.9	4.9	4.4	5.9	3.3	4.1	6.3	7.4
		0.4	6.5	30.8	24.9	31.3	25.5	6.8	18.7	23.5	20.1	6.5	14.3
		0.8	10.0	69.9	60.0	67.6	61.1	32.8	55.8	62.0	57.8	15.3	48.1
	256	0.0	4.5	4.4	3.7	6.8	5.2	5.9	4.2	5.1	4.8	5.5	4.7
		0.4	7.1	67.0	55.2	66.3	53.7	11.1	35.6	54.8	42.5	6.0	24.0
		0.8	7.9	97.8	93.5	97.1	92.9	74.3	91.5	95.3	91.2	41.0	85.0
	512	0.0	4.8	4.7	4.6	8.6	7.5	4.7	4.9	7.4	6.9	4.4	4.6
		0.4	6.5	96.0	90.8	94.3	88.5	22.4	75.5	89.4	83.1	8.2	51.6
		0.8	7.2	100.0	99.7	100.0	99.6	99.3	99.6	99.9	99.2	89.2	99.6

Table 8: Approximate p-values (multiplied by 100) of the rank-based tests of stationarity proposed in Section 2 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 meaning of the abbreviations d, c, dc and dcp is given in Section 5. The columns c2 and c3 report the results for the bivariate analogues of the test based on $S_n^{(2)}$ defined by (2.3) (which arise in the combined test dcp) for lags 2 and 3.

		$h = 2$ 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