Submitted to Bernoulli

A bootstrapped test of covariance stationarity based on orthonormal transformations

Jonathan B. Hill[1] and Tianqi Li[2]

[1]*Dept. of Economics, University of North Carolina*
[2]*Dept. of Economics, University of North Carolina*

We propose a covariance stationarity test for an otherwise dependent and possibly globally non-stationary time series. We work in a generalized version of the new setting in Jin, Wang and Wang (2015), who exploit Walsh (1923) functions in order to compare sub-sample covariances with the full sample counterpart. They impose strict stationarity under the null, only consider linear processes under either hypothesis in order to achieve a parametric estimator for an inverted high dimensional asymptotic covariance matrix, and do not consider any other orthonormal basis. Conversely, we work with a general orthonormal basis under mild conditions that include Haar wavelet and Walsh functions; and we allow for linear or nonlinear processes with possibly non-iid innovations. This is important in macroeconomics and finance where nonlinear feedback and random volatility occur in many settings. We completely sidestep asymptotic covariance matrix estimation and inversion by bootstrapping a max-correlation difference statistic, where the maximum is taken over the correlation lag h and basis generated sub-sample counter k (the number of systematic samples). We achieve a higher feasible rate of increase for the maximum lag and counter $\mathcal{H}_T$ and $\mathcal{K}_T$. Of particular note, our test is capable of detecting breaks in variance, and distant, or very mild, deviations from stationarity.

Keywords: covariance stationarity; max-correlation test; multiplier bootstrap; orthonormal basis; Walsh functions

1. Introduction

Assume $\{X_t : t \in \mathbb{Z}\}$ is a possibly non-stationary time series process in $\mathcal{L}_2$. We want to test whether X_t is covariance stationary, without explicitly assuming stationarity under the null hypothesis, allowing for linear or nonlinear processes with a possibly non-iid innovation, and a general memory property. Such generality is important in macroeconomics and finance where nonlinear feedback and non-iid innovations occur in many settings due to asymmetries and random volatility, including exchange rates, bonds, interest rates, commodities, and asset return levels and volatility. Popular models for such time series include symmetric and asymmetric GARCH, Stochastic Volatility, nonlinear ARMA-GARCH, and switching models like smooth transition autoregression. See, e.g., Teräsvirtra (1994), Gray (1996) and Francq and Zakoïan (2019).

Evidence for nonstationarity, whether generally or in the variance or autocovariances, has been suggested for many economic time series, where breaks in variance and model parameters are well known (e.g. Busett and Taylor, 2003, Gianetto and Raissi, 2015, Hendry and Massmann, 2007, Perron, 2006). Knowing whether a time series is globally nonstationary has large implications for how analysts approach estimation and inference. Indeed, it effects whether conventional parametric and semi-(non)parametric model specifications are correct. Pretesting for deviations from global stationarity therefore has important practical value.

There are many tests in the literature on covariance stationarity, and concerning locally stationary processes. Tests for stationarity based on spectral or second order dependence properties have a long history, where pioneering work is due to Priestley and Subba Rao (1969). Spectrum-based tests with $\mathcal{L}_2$-distance components have many versions. Paparoditis (2010a) uses a rolling window method to compare subsample local periodograms against a full sample version. The maximum is taken over the $\mathcal{L}_2$-distance between periodograms over all time points. An asymptotic theory for the max-statistic, however, is not provided, although an approximation theory is (see their Lemmas 1 and 3). Furthermore, conforming with many offerings in the literature, under the null X_t is a linear process with iid Gaussian innovations. Dette, Preuß and Vetter (2011) study locally stationary processes, and impose linearity with iid Gaussian innovations. Their statistic is based on the minimum $\mathcal{L}_2$-distance between a spectral density and its version under stationarity, and local power is non-trivial against $T^{1/4}$-alternatives. Aue et al. (2009) propose a nonparametric test for break in covariance for multivariate time series based on a version of a cumulative sum statistic.

Wavelet methods have arisen in various forms recently. von Sachs and Neumann (2000), using technical wavelet decomposition components from Neumann and von Sachs (1997), propose a Haar wavelet based localized periodogram test of covariance stationarity for locally stationary processes (cf. Dahlhaus, 1997, 2009). Local and asymptotic power are not theoretically derived. Haar wavelet functions form an orthonormal basis on $\mathcal{L}_2[0, 1)$, but the proposed frequency domain tests are complicated, a local power analysis is not feasible, and empirical power may be weak (see Jin, Wang and Wang's (2015) simulation evidence).

Dwivedi and Subba Rao (2011) and Jentsch and Rao (2015) use the discrete Fourier transform [DFT] $J_T(\omega_k) = (2\pi T)^{-1/2}\sum_{t=1}^{T} X_t \exp\{it\omega_k\}$ at canonical frequencies $\omega_k = 2\pi k/T$ and $1 \leq k \leq T$. Dwivedi and Subba Rao (2011) generate a portmanteau statistic from a normalized sample DFT covariance, exploiting the fact that an uncorrelated DFT implies second order stationarity. Nason (2013) presents a covariance stationarity test based on Haar wavelet coefficients of the wavelet periodogram, they assume linear local stationarity, and do not treat local power. See also Nason, von Sachs and Kroisandt (200).

In a promising offering in the wavelet literature, Jin, Wang and Wang (2015) [JWW] exploit so-called *Walsh functions* (akin to "global square waves" although not truly wavelets; cf. Walsh 1923) and their implied systematic samples for comparing sub-sample covariances with the full sample one. They utilize a sample-size dependent maximum lag $\mathcal{H}_T$ and maximum systematic sample counter $\mathcal{K}_T$, and show their Wald test exhibits non-negligible local power against $\sqrt{T}$-alternatives. They do not consider any other orthonormal transformation because Walsh functions, they argue, have "*desirable properties*" based primarily on simulation evidence, asymptotic independence of a sub-sample and sample covariance difference $(\sqrt{T}(\hat{\gamma}_h^{(k_1)} - \hat{\gamma}_h), \sqrt{T}(\hat{\gamma}_h^{(k_2)} - \hat{\gamma}_h))$ across systematic samples $k_1 \neq k_2$, and joint asymptotic normality (JWW, p. 897). It seems, however, that such theoretical properties are available irrespective of the orthonormal basis used, although we do not provide a proof. See Section 2.1, below, for definitions and notation. We do, however, find in the sequel that the Walsh basis has superlative properties vis-à-vis a Haar wavelet basis.

JWW's asymptotic analysis is driven by local stationarity and linearity $X_t = \sum_{i=0}^{\infty}\psi_i Z_{t-i}$, with zero mean iid Z_t, and $E|Z_t|^{4+\delta} < \infty$, $\delta > 0$, which expedites characterizing a parametric asymptotic covariance matrix estimator. The iid and linearity assumptions, however, rule out many important processes, including nonlinear models like regime switching, and random coefficient processes; and any process with a non-iid error (e.g. nonlinear ARMA-GARCH). JWW's Wald-type test statistic requires an inverted parametric variance estimator that itself requires five tuning parameters and choice of two kernels.[1] Indeed, most of the tuning parameters only make sense under linearity given how they approach asymptotic covariance matrix estimation.

[1] One tuning parameter $\lambda \in (0, .5)$ governs the number $Q_T = [T^{\lambda}]$ of sample covariances that enter the asymptote covariance matrix estimator (see their p. 899); and four $(c_1, c_2; \xi_1, \xi_2)$ are used for kernel bandwidths $b_j = c_j T^{-\xi_j}$, $j = 1, 2$, for computing the kurtosis of the iid process Z_t under linearity (see p. 902-903). The authors set c_j equal to 1.2 times a so-called "*crude scale estimate*" which is nowhere defined.

Now define the lag h autocovariance coefficient at time t:

$$\gamma_h(t) \equiv E\left[(X_t - E[X_t])(X_{t-h} - E[X_{t-h}])\right], h = 0, 1, \ldots$$

The hypotheses are:

$$H_0 : \gamma_h(s) = \gamma_h(t) = \gamma_h \ \forall s, t, \forall h = 0, 1, \ldots \text{ (cov. stationary)} \tag{1.1}$$
$$H_1 : \gamma_h(s) \neq \gamma_h(t) \text{ for some } s \neq t \text{ and } h = 0, 1, \ldots \text{ (cov. nonstationary).}$$

Under H_0 X_t is second order stationary, and the alternative is *any* deviation from the null: the autocovariance differs across time at some lag, allowing for a (lag zero) break in variance. The null hypothesis otherwise accepts the possibility of global nonstationarity.

In this paper we do away with parametric assumptions on X_t, and impose a mixing property that allows us to bound the number of usable covariance lags $\mathcal{H}_T$ and systematic samples $\mathcal{K}_T$. The mixing condition allows for global nonstationarity under either hypothesis, allowing us to focus the null hypothesis only on second order stationarity.

Rather than operate on a Wald statistic constructed from a specific orthonormal transformation of covariances, our statistic is the maximum generic orthonormal transformed sample correlation coefficient, where the maximum is taken over (h, k) with increasing upper bounds $(\mathcal{H}_T, \mathcal{K}_T)$. Notice h is the covariance lag, and k is a counter for a particular systematic subsample implied by the specific transformation. By working in a generic setting we are able to make direct comparisons, and combine bases for possible power improvements.

We provide examples of Haar wavelet and Walsh functions in Sections 2.1 and 2.2, and show how they yield different systematic samples. This suggests a power improvement may be available by using *multiple* orthonormal transforms. As JWW (p. 897) note, however, clearly other orthonormal transformations are feasible, although simulation evidence agrees with their suggestion that the Walsh basis works quite well.

We use a dependent wild bootstrap for the resulting test statistic, which allows us to sidestep asymptotic covariance matrix estimation, a challenge considering we do not assume a parametric form, and the null hypothesis requires us to look over a large set of (h, k). We sidestep all of JWW's tuning parameters, and require just one governing the block size for the bootstrap. We ultimately achieve a significantly better upper bound on the rate of increase for $(\mathcal{H}_T, \mathcal{K}_T)$ than JWW. We show that penalized and weighted versions of our test statistic are possible, as in JWW and Hill and Motegi (2020) respectively. However, we argue that there is no compelling theory to justify penalties on (h, k) in our setting, and overall a non-penalized and unweighted test statistic works best in practice.

Note that Hill and Motegi (2020) study the max-correlation statistic for a white noise test, and only show their limit theory applies for some increasing maximum lag $\mathcal{H}_T$, but do not derive an upper bound. In the present paper we use a different asymptotic theory, derive upper bounds for $\mathcal{H}_T$ and $\mathcal{K}_T$, and of course do not require a white noise property under H_0.

Jin, Wang and Wang (2015, Section 2.6) rule out the use of autocorrelations because, they claim, if the sample variance were included, i.e. $h \geq 0$, then consistency may still not hold because the limit theory neglects the joint distribution of $\hat{\gamma}_0$ and the correlation differences. We show for our proposed test that the difference between full sample and systematic sample autocorrelations at lag zero asymptotically reveals whether $E[X_t^2]$ is time dependent. Further, our test is consistent whether non-stationarity is caused by variances, or covariances, or both. See Section 3.1, and Example 3.5. Our proposed test is consistent against a general (nonparametric) alternative, and exhibits nontrivial power against a sequence of $\sqrt{T}$-local alternatives.

Finally, in the supplemental material we present an automatic method for selecting $\mathcal{H}_T$ and $\mathcal{K}_T$. The method is based on Hill and Motegi's (2020) extension of a maximum lag selection technique for a Q-test based portmanteau statistic in Escanciano and Lobato (2009), cf. Inglot and Ledwina (2006). It requires sets $\{0, ..., \bar{\mathcal{H}}_T\}$ and $\{1, ..., \bar{\mathcal{K}}_T\}$ from which optimal values $\{\mathcal{H}_T^*, \mathcal{K}_T^*\}$ are iteratively selected, hence we still need pre-chosen maxima $\bar{\mathcal{H}}_T$ and $\bar{\mathcal{K}}_T$. In simulation experiments not reported here, however, we find the Escanciano and Lobato (2009) logic applied to an orthonormal transformed sample covariance does not lead to a dominant test the way it can for a max-correlation white noise test (cf. Hill and Motegi, 2020). The method systematically sets $\{\mathcal{H}_T^*, \mathcal{K}_T^*\}$ to the lowest value capable of detecting a deviation from covariance stationarity. This generally leads to (very) small values of $\{\mathcal{H}_T^*, \mathcal{K}_T^*\}$ and therefore low empirical power. Merely using a pre-chosen $\{\mathcal{H}_T, \mathcal{K}_T\}$ leads to sharp size and competitive power. Whether another data-dependent method applies is left for future work.

The max-correlation difference is particularly adept at revealing subtle deviations from covariance stationarity, similar to results revealed in Hill and Motegi (2020). Consider a distant form of a model treated in Paparoditis (2010b, Model I) and Jin, Wang and Wang (2015, Section 3.2: models NVI, NVII), $X_t = .08\cos\{1.5 - \cos(4\pi t/T)\}\epsilon_{t-d} + \epsilon_t$ with large d (JWW use $d = 1$ or 6). JWW's test exhibits trivial power when $d \geq 20$, while the max-correlation difference is able to detect this deviation from the null even when $d \geq 50$. The reason is the same as that provided in Hill and Motegi (2020): the max-correlation difference operates on the single most useful statistic, while Wald and portmanteau statistics congregate many standardized covariances that generally provide little relevant information under a weak signal.

In Section 2 we develop the test statistic. Sections 3 and 4 present asymptotic theory and the bootstrap method and theory. We then perform a Monte Carlo study in Section 5, and conclude with Section 6. Proofs are presented in Appendix A. The supplemental material contains omitted proofs, a data-dependent method for selecting $(\mathcal{H}_T, \mathcal{K}_T)$, an empirical study concerning international interest rates, and complete simulation results.

We use the following notation. $[z]$ rounds z to the nearest integer.. $\mathcal{L}_2$ is the space of square integrable random variables; and $\mathcal{L}_2[a, b)$ is the class of square integrable functions on $[a, b)$. $||\cdot||_p$ and $||\cdot||$ are the L_p and l_2 norms respectively, $p \geq 1$. Let $\mathbb{Z} \equiv \{... -2, -1, 0, 1, 2, ...\}$, and $\mathbb{N} \equiv \{0, 1, 2, ...\}$. $K > 0$ is a finite constant whose value may be different in different places. $awp1$ denotes "*asymptotically with probability approaching one*". Write $\max_{\mathcal{H}_T} = \max_{0 \leq h \leq \mathcal{H}_T}$. $\max_{\mathcal{K}_T} = \max_{1 \leq k \leq \mathcal{K}_T}$ and $\max_{\mathcal{H}_T, \mathcal{K}_T} = \max_{0 \leq h \leq \mathcal{H}_T, 1 \leq k \leq \mathcal{K}_T}$. Similarly, $\max_{\mathcal{H}_T} a(h, \tilde{h}) = \max_{0 \leq h, \tilde{h} \leq \mathcal{H}_T} a(h, \tilde{h})$, etc.

2. Max-correlation with orthonormal transformation

Our test statistic is the maximum of an orthonormal transformed sample covariance. In order to build intuition, we first derive the test statistic under Walsh function and Haar wavelet-based bases. We then set up a general environment, and present the main results.

In order to reduce notation, we assume here $\mu \equiv E[X_t] = 0$ is known. In practice this is enforced by using $X_t - \bar{X}$ where $\bar{X} \equiv 1/T \sum_{t=1}^T X_t$. In the appendix we prove using $X_t - \bar{X}$ or $X_t - \mu$ leads to identical results asymptotically: see Lemma A.3. Thus in proofs of the main results we simply assume $\mu = 0$.

2.1. Walsh functions

The following class of Walsh functions $\{W_i(x)\} \equiv \{W_i(x) : i = 0, 1, 2, ...\}$ define a complete orthonormal basis in $\mathcal{L}_2[0, 1)$. The functions $W_i(x)$ are defined recursively (see, e.g., Ahmed and Rao, 1975, Stoffer, 1987, 1991, Walsh, 1923):

$$W_0(x) = 1 \text{ for } x \in [0, 1); \text{ and } W_1(x) = \begin{cases} 1, & x \in [0, .5) \\ -1, & x \in [.5, 1) \end{cases}$$

and for any $i = 1, 2, ...,$

$$W_{2i}(x) = \begin{cases} W_i(2x), & x \in [0, .5) \\ (-1)^i W_i(2x-1), & x \in [.5, 1) \end{cases} \text{ and } W_{2i+1}(x) = \begin{cases} W_i(2x), & x \in [0, .5) \\ (-1)^{i+1} W_i(2x-1), & x \in [.5, 1) \end{cases}.$$

In the $\{-1, 1\}$-valued sequence $\{W_i(x) : i = 0, 1, 2, ...\}$, i indexes the number of zero crossings, yielding a square shaped wave-form. See Figure 1, and see Stoffer (1991, Figure 5) and Jin, Wang and Wang (2015, Figure 1) and their references. The k^{th} discrete Walsh functions used in this paper are then for $t = 1, ..., T$:

$$\{\mathcal{W}_k(1), ..., \mathcal{W}_k(T)\} \text{ where } \mathcal{W}_k(t) = W_k((t-1)/T).$$

Now define the covariance coefficient for a covariance stationary time series, $\gamma_h \equiv E[X_t X_{t-h}]$; and denote the usual (co)variance estimator $\hat{\gamma}_h \equiv 1/T \sum_{t=1}^{T-h} X_t X_{t+h}$, $h \in \mathbb{N}$. JWW use $\{\mathcal{W}_i(x)\}$ to construct a set of discrete Walsh covariance transformations:

$$\hat{\gamma}_h^{W(k)} \equiv \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} \left\{1 + (-1)^{k-1} \mathcal{W}_k(t)\right\}, \ h = 0, 1, ..., T-1, \text{ and } k = 1, 2, ..., \mathcal{K} \tag{2.1}$$

for some integer $\mathcal{K} \geq 1$. As they point out, a sequence of systematic (sub)samples $\boldsymbol{T}_k^W : k = 1, 2, ..., \mathcal{K}$ in the time domain can be defined on the basis of Walsh functions:

$$\boldsymbol{T}_k^W \equiv \left\{t \in T : (-1)^{k-1} \mathcal{W}_k(t) = 1\right\}.$$

Now let $\mathcal{N}_k$ be the smallest power of 2 that is at least k. The first systematic sample is the first half of the sample time domain $\boldsymbol{T}_1^W = \{1, ..., [T/2]\}$; the second is the middle half $\boldsymbol{T}_2^W = \{[T/4], [T/4] + 1, ..., [3T/4]\}$; the third $\boldsymbol{T}_3$ is the first and third time blocks, and so on. Notice $\boldsymbol{T}_k^W$ consists of $(k+1)/2$ blocks with at least $[T/\mathcal{N}_k]$ elements. Thus, when $h < T/\mathcal{N}_k$ then $\hat{\gamma}_h^{W(k)}$ is just an estimate of γ_h on the k^{th} systematic sample:

$$\hat{\gamma}_h^{W(k)} = \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} \left\{1 + (-1)^{k-1} \mathcal{W}_k(t)\right\} = \frac{2}{T} \sum_{t \in \boldsymbol{T}_k} X_t X_{t+h}.$$

The condition $h < T/\mathcal{N}_k$ holds asymptotically in the Section 4 bootstrap setting.

The difference between the k^{th} systematic sample and full sample estimators is:

$$\hat{\gamma}_h^{W(k)} - \hat{\gamma}_h = (-1)^{k-1} \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} \mathcal{W}_k(t).$$

Notice the $\{-1, 1\}$-valued nature of $\mathcal{W}_k(t)$ yields a sub-sample comparison: $\hat{\gamma}_h^{W(k)} - \hat{\gamma}_h = 1/T \sum_{t \in \boldsymbol{T}_k} X_t X_{t+h} - 1/T \sum_{t \notin \boldsymbol{T}_k} X_t X_{t+h}$. Our test is based on the maximum $|\hat{\gamma}_h^{W(k)} - \hat{\gamma}_h|$, in which case the multiple $(-1)^{k-1}$ is irrelevant. We, therefore, now drop it everywhere. Under the null hypothesis and mild assumptions this difference is $O_p(1/\sqrt{T})$ at all lags h and for all systematic samples k. Thus, a test statistic can be constructed from $\sqrt{T}(\hat{\gamma}_h^{W(k)} - \hat{\gamma}_h)$.

2.2. Haar wavelet functions

Define the usual Haar wavelet functions $\psi_{k,m}(x) \equiv 2^{k/2}\psi(2^k x - m)$ with $x \in \mathbb{R}$, where $0 \leq k \leq \mathcal{K}_T$ for some integer sequence $\{\mathcal{K}_T\}$, $0 \leq m \leq 2^k - 1$, and mother wavelet (Haar, 1910):

$$\psi(x) = \begin{cases} 1, & x \in [0, .5) \\ -1, & x \in [.5, 1) \\ 0 & otherwise \end{cases}$$

Haar functions $\{\psi_{k,m}(x)\}$ form a complete orthonormal basis in $\mathcal{L}[0,1)$. The discretized version is:

$$\Psi_{k,m}(t) \equiv \psi_{k,m}((t-1)/T) = 2^{k/2}\psi(2^k (t-1)/T - m).$$

Systematic samples derived from $\{\Psi_{k,m}(t)\}$ are generally too "local": $1/T\sum_{t=1}^{T-h} X_t X_{t+h}\Psi_{2,m}(t)$, for example, compares just the first eighth to the second eighth subsample ($m = 0$); the third eighth to the fourth eighth subsample ($m = 1$); and so on.

In order to yield a test statistic that compares sub-sample complements, comparable to Walsh functions, we compile $(\psi_{k,m}(x), \Psi_{k,m}(t))$ over $0 \leq m \leq 2^k - 1$. Set $\psi_0(x) = I(0 \leq x \leq 1)$, and for $k = 0, 1, \ldots$

$$\psi_{k+1}(x) \equiv \frac{1}{2^{k/2}}\sum_{m=0}^{2^k-1}\psi_{k,m}(x) = \sum_{m=0}^{2^k-1}\psi(2^k x - m)$$

$$\Psi_{k+1}(t) \equiv \frac{1}{2^{k/2}}\sum_{m=0}^{2^k-1}\Psi_{k,m}(t) = \sum_{m=0}^{2^k-1}\psi(2^k (t-1)/T - m).$$

We set $\psi_0(x) = I(0 \leq x \leq 1)$ in order to unify the local alternative analysis below, similar to the Walsh basis. It can be shown that $\psi_k(x) \in \{-1, 1\}$, and $\{\psi_k(x) : 1 \leq k \leq \mathcal{K}_T\}$ forms a complete orthonormal basis: see Lemma A.1 in the appendix for this and other properties. In the same manner as (2.1), define for $k = 1, 2, \ldots, \mathcal{K}$:

$$\hat{\gamma}_h^{H(k)} \equiv \frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h}\{1 + \Psi_k(t)\}, \ h = 0, 1, \ldots, T-1.$$

The discretized Haar functions $\Psi_{k,m}(t)$ generate systematic samples $\boldsymbol{T}_k^H \equiv \{t \in T : \bigcup_{m=0}^{2^k-1}(\Psi_{k,m}(t) = 1)\}$. This yields the first sample half $\boldsymbol{T}_0^H = \{1, \ldots, [T/2]\}$; the first and third quarter subsamples $\boldsymbol{T}_1^H = \{1, \ldots, [T/4]; 1 + [T/2], \ldots, [3T/4]\}$; the first, third, fifth and seventh eights $\boldsymbol{T}_2^H$; and so on. See Figures 1 and 2 for plots of Walsh and composite Haar functions $W_k(x)$ and $\psi_k(x)$, $k = 1, \ldots, 6$.

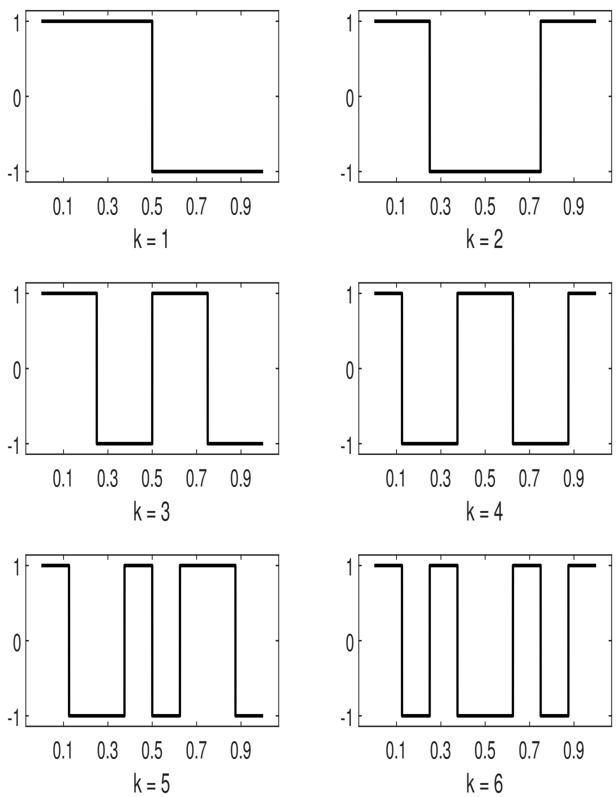

Figure 1: Walsh functions $\{W_k(x)\}_{k=1}^{6}$.

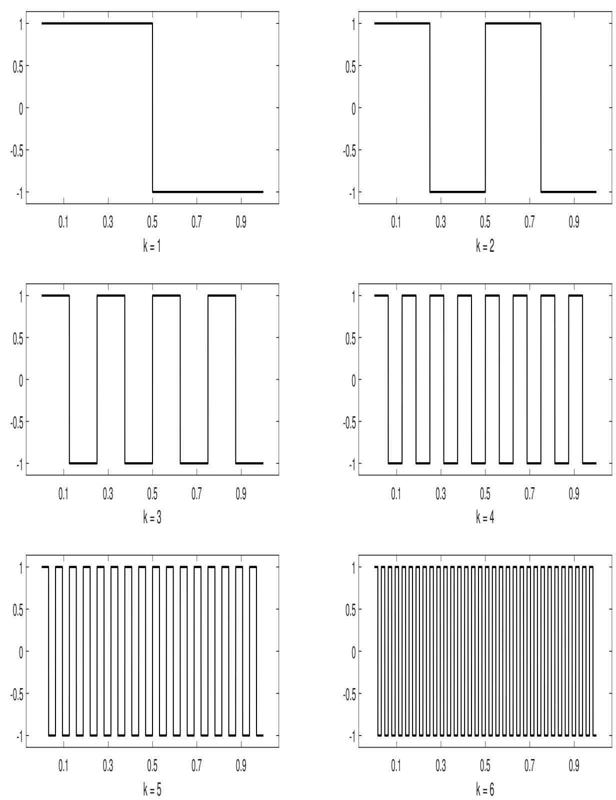

Figure 2: Composite Haar wavelet functions $\{\psi_k(x)\}_{k=1}^{6}$

Walsh and Haar systematic samples are quite different for $k \geq 2$. $\boldsymbol{T}_k^W$ involves fewer interspersed subsample segments, in some cases of varying lengths, while $\boldsymbol{T}_k^H$ have 2^k segments of equal length $[T/2^k]$ in all cases (ignoring truncation due to the lag h). Haar subsamples are therefore non-redundant only when $T/2^{\mathcal{K}_T} \geq 1$, hence $\mathcal{K}_T \leq \ln(T)/\ln(2)$.

Indeed, it can be shown that the two bases coincide in the sense that $W_{k_1}(x) = \psi_{k_2}(x)$ for all x and only $(k_1, k_2) \in \{(1,1),(3,2),(3,7)\}$, or in all other cases for x on a subset of $[0,1)$ with Lebesgue measure $1/2$. Roughly speaking, only 50% of the data points in $\hat{\gamma}_h^{W(k_1)} - \hat{\gamma}_h$ are the same as those in $\hat{\gamma}_h^{H(k_2)} - \hat{\gamma}_h$ for nearly all systematic samples (k_1, k_2). Thus the two bases are intrinsically different, suggesting potential advantages and weaknesses against certain deviations from the null.

2.3. Max-correlation orthonormal transforms

Now let $\{\mathcal{B}_k(x) : 0 \leq k \leq \mathcal{K}\}$ denote a $\{-1,1\}$-valued orthonormal basis on $\mathcal{L}[0,1)$, $\mathcal{B}_0(x) = I(0 \leq 0 \leq 1)$, let $B_k(t) \equiv \mathcal{B}_k((t-1)/T)$ be the discretized version, and define the usual subsample covariance for this generic discrete basis $\hat{\gamma}_h^{(k)} \equiv 1/T \sum_{t=1}^{T-h} X_t X_{t+h}\{1 + B_k(t)\}$. Under conditions imposed below, examples of $B_k(t)$ are Walsh $\mathcal{W}_k(t)$ and Haar composite $\Psi_k(t)$.

Define the sample correlation coefficient:

$$\hat{\rho}_h \equiv \frac{\hat{\gamma}_h}{\hat{\gamma}_0},$$

and a set of discrete orthonormal correlation transformations, over systematic sample k:

$$\hat{\rho}_{h,1}^{(k)} \equiv \frac{\hat{\gamma}_h^{(k)}}{\hat{\gamma}_0} = \frac{1}{\hat{\gamma}_0} \times \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} \{1 + B_k(t)\}, \ k = 1, 2, ..., \mathcal{K}.$$

Thus, the difference between systematic sample and full sample estimators is:

$$\hat{\rho}_{h,1}^{(k)} - \hat{\rho}_h = \frac{1}{\hat{\gamma}_0} \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} B_k(t) = \frac{\hat{\gamma}_h^{(k)} - \hat{\gamma}_h}{\hat{\gamma}_0} \tag{2.2}$$

The correlation difference $\hat{\rho}_{h,1}^{(k)} - \hat{\rho}_h$ is sensible even at lag 0, considering

$$\hat{\rho}_{0,1}^{(k)} - \hat{\rho}_0 = \frac{\hat{\gamma}_0^{(k)}}{\hat{\gamma}_0} - 1.$$

Thus, under nonstationarity $\hat{\rho}_{0,1}^{(k)} \overset{p}{\nrightarrow} 1$ for some systematic sample k when $\hat{\gamma}_0^{(k)}/\hat{\gamma}_0 \overset{p}{\nrightarrow} 1$; that is, when the second moment $E[X_t^2]$ is not constant over t.

Alternatively, we may incorporate the systematic sample variance estimators $\hat{\gamma}_0^{(k)}$. The autocorrelation estimator in that case becomes, for example:

$$\begin{aligned}
\hat{\rho}_{h,2}^{(k)} &\equiv \frac{1}{\hat{\gamma}_0} \times \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} + \frac{1}{\hat{\gamma}_0^{(k)}} \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} B_k(t) \\
&= \hat{\rho}_h + \frac{1}{\hat{\gamma}_0^{(k)}} \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} B_k(t)
\end{aligned}$$

hence

$$\hat{\rho}_{h,2}^{(k)} - \hat{\rho}_h = \frac{\frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h} B_k(t)}{\frac{1}{T}\sum_{t=1}^{T} X_t^2 \{1 + B_k(t)\}}.$$

At lag 0 notice:

$$\hat{\rho}_{0,2}^{(k)} - \hat{\rho}_0 = \frac{1}{\hat{\gamma}_0^{(k)}} \frac{1}{T} \sum_{t=1}^{T} X_t^2 B_k(t) = \frac{\hat{\gamma}_0^{(k)} - \hat{\gamma}_0}{\hat{\gamma}_0^{(k)}} = 1 - \frac{\hat{\gamma}_0}{\hat{\gamma}_0^{(k)}}.$$

Compare this to $\hat{\rho}_{0,1}^{(k)} - \hat{\rho}_0 = \hat{\gamma}_0^{(k)}/\hat{\gamma}_0 - 1$. Thus, again $\hat{\rho}_{0,2}^{(k)} \overset{p}{\nrightarrow} 1$ for some systematic sample k when $E[X_t^2]$ is not constant over t.

The autocorrelation estimators $\hat{\rho}_{h,i}^{(k)}$ exploit orthonormal transform weights $B_k(t)$ in order to reveal autocorrelation subsample differences, but they are not identical in small samples. Asymptotically, however, their difference is negligible in probability under the null hypothesis. Notice:

$$\begin{aligned}\hat{\rho}_{h,1}^{(k)} - \hat{\rho}_{h,2}^{(k)} &= \left(\frac{1}{\hat{\gamma}_0} - \frac{1}{\hat{\gamma}_0^{(k)}}\right) \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} B_k(t) \\ &= \left(\hat{\gamma}_0^{(k)} - \hat{\gamma}_0\right) \frac{1}{\hat{\gamma}_0 \hat{\gamma}_0^{(k)}} \left(\hat{\gamma}_h^{(k)} - \hat{\gamma}_h\right). \end{aligned} \tag{2.3}$$

This reveals $\hat{\rho}_{h,1}^{(k)} - \hat{\rho}_{h,2}^{(k)}$ for each $h \geq 0$ simultaneously captures systematic sample differences in variance *and* covariance. Under H_0 and general conditions presented in Section 3, $\max_{\mathcal{H}_T,\mathcal{K}_T} |\hat{\gamma}_h^{(k)} - \hat{\gamma}_h|$ and $|\hat{\gamma}_0 - \gamma_0|$ are $O_p(1/\sqrt{T})$, where $\{\mathcal{H}_T, \mathcal{K}_T\}$ are sequences defined below with $\mathcal{H}_T \to \infty$ and $\mathcal{K}_T \to \infty$. Thus:

$$\max_{\mathcal{H}_T,\mathcal{K}_T} \left|\sqrt{T}\left(\hat{\rho}_{h,1}^{(k)} - \hat{\rho}_{h,2}^{(k)}\right)\right| = O_p(1/\sqrt{T}).$$

Under H_1, however, *if and only if* $E[X_t^2]$ and $E[X_t X_{t-h}]$ *for some* $h \geq 1$ are time dependent then $\sqrt{T} \max_{\mathcal{H}_T,\mathcal{K}_T} |\hat{\rho}_{h,1}^{(k)} - \hat{\rho}_{h,2}^{(k)}| \overset{p}{\to} \infty$. This suggests $\mathcal{D}_T \equiv \max_{\mathcal{H}_T,\mathcal{K}_T} |\sqrt{T}(\hat{\rho}_{h,1}^{(k)} - \hat{\rho}_{h,2}^{(k)})|$ could be used as a third test statistic: theory developed in Section 3 can be used to show a test based on $\mathcal{D}_T$ will reject H_0 asymptotically with power approaching one when X_t is non-stationary in variance *and* autocovariance at some lag $h \geq 1$. Conversely, either $\max_{\mathcal{H}_T,\mathcal{K}_T} |\sqrt{T}(\hat{\rho}_{h,i}^{(k)} - \hat{\rho}_h)|$ is consistent against H_1 in general: power is one asymptotically if $E[X_t^2]$ *and/or* some $E[X_t X_{t-h}]$ are time dependent.

In order to focus ideas, however, we only consider the estimator $\hat{\rho}_{h,1}^{(k)}$, so put:

$$\hat{\rho}_h^{(k)} \equiv \hat{\rho}_{h,1}^{(k)}.$$

The proposed test statistic is therefore the maximum normalized $\hat{\rho}_h^{(k)} - \hat{\rho}_h$ over (h, k):

$$\mathcal{M}_T \equiv \sqrt{T} \max_{\mathcal{H}_T,\mathcal{K}_T} \left|\hat{\rho}_h^{(k)} - \hat{\rho}_h\right| = \frac{1}{\hat{\gamma}_0} \max_{\mathcal{H}_T,\mathcal{K}_T} \left|\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} X_t X_{t+h} B_k(t)\right|.$$

By construction $\mathcal{M}_T$ uses the most informative systematic sample correlation difference. Notice we search over *all* lags $h \in \{0, ..., \mathcal{H}_T\}$.

A penalized version is also possible:

$$\mathcal{M}_T^{(p)} \equiv \max_{\mathcal{H}_T,\mathcal{K}_T} \left\{ \sqrt{T} \left| \hat{\rho}_h^{(k)} - \hat{\rho}_h \right| - \mathcal{P}(h,k) \right\},$$

where $\mathcal{P}(h,k)$ is a non-random, positive, strictly monotonically increasing function of h and k. JWW use the additive $\mathcal{P}(h,k) = p_h + q_h$ with AIC-like lag penalty $p_h = 2h$ in an order selection-type Wald statistic. This is sensible considering the Wald statistic is pointwise asymptotically chi-squared with mean $2h$ for each k (see also Inglot and Ledwina, 2006). For q_k they use $\sqrt{k-1}$ based primarily on empirical power considerations.[2]

In our non-Wald setting a similar reasoning for choosing $\mathcal{P}(h,k) = p_h + q_h$ does not apply, nor do we have any comparable requirements for penalizing k. Indeed, a compelling reason for "penalizing" $\mathcal{M}_T$ at all would be to counter the loss of observations at higher lags or to control for lag specific heterogeneity, but that historically is ameliorated with a weighted correlation, for example $\max_{\mathcal{H}_T,\mathcal{K}_T}\{\sqrt{T}\mathfrak{W}_{T,h}^{(k)}|\hat{\rho}_h^{(k)} - \hat{\rho}_h|\}$, where $\mathfrak{W}_{T,h}^{(k)}$ are possibly stochastic weights, $\liminf_{T\to\infty} \min_{\mathcal{H}_T,\mathcal{K}_T} \mathfrak{W}_{T,h}^{(k)} > 0$ $a.s.$, and $\max_{\mathcal{H}_T,\mathcal{K}_T} |\mathfrak{W}_{T,h}^{(k)} - \mathfrak{W}_h^{(k)}| \xrightarrow{p} 0$ where the non-stochastic $\mathfrak{W}_h^{(k)}$ satisfy $\min_{\mathcal{H}_T,\mathcal{K}_T} \mathfrak{W}_h^{(k)} > 0$. Choices include Ljung-Box type weights, or an inverted nonparametric standard deviation estimator, cf. Hill and Motegi (2020).

Consider the latter, define a sample covariance function $\hat{v}_T(i;h,k) \equiv 1/T \sum_{t=1}^{T-h-i} \hat{z}_t(h,k)\hat{z}_{t+i}(h,k)$ where

$$\hat{z}_t(h,k) \equiv \left\{ X_t X_{t+h} B_k(t) - \frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h} B_k(t) \right\}.$$

Under fourth order stationarity under the null, and because $\hat{\gamma}_0$ only operates as a scale asymptotically, cf. Theorem 3.3 below, the weights are $\mathfrak{W}_{T,h}^{(k)} = 1/\hat{\mathcal{V}}_T(h,k)$ where, e.g.,

$$\hat{\mathcal{V}}_T^2(h,k) = \hat{\gamma}_0^{-2} \left\{ \hat{v}_T(0;h,k) + 2\sum_{i=1}^{T-h-1} \mathcal{K}(i/\beta_T)\hat{v}_T(i;h,k) \right\} \tag{2.4}$$

with symmetric, square integrable kernel function $\mathcal{K} : \mathbb{R} \to [-1,1]$ satisfying $\mathcal{K}(0) = 1$,[3] and bandwidth $\beta_T \to \infty$ and $\beta_T = o(T)$.

A penalized and weighted version is thus:

$$\mathcal{M}_T^{(w,p)} \equiv \max_{0\le h\le\mathcal{H}_T, 1\le k\le\mathcal{K}_T} \left\{ \sqrt{T}\mathfrak{W}_{T,h}^{(k)} \left| \hat{\rho}_h^{(k)} - \hat{\rho}_h \right| - \mathcal{P}(h,k) \right\}. \tag{2.5}$$

In Monte Carlo work we study $\mathcal{M}_T$, $\mathcal{M}_T^{(p)}$, and $\mathcal{M}_T^{(w,p)}$ with a Walsh or Haar basis, and various penalties and/or an inverted standard deviation weight or Ljung-Box weight. We find using $\mathcal{P}(h,k) = p_h + q_k$ where $p_h = (h+1)^a/2$ and $q_k = k^a/2$ with $a = [1/8, 1/2]$, or $\mathcal{P}(h,k) = \sqrt{(h+1)k}$, promotes accurate empirical size but generally does not lead to dominant power, and may lead to decreased power in some cases. Conversely, stochastic weights $\mathfrak{W}_{T,h}^{(k)}$ generally lead to over-sized tests, while Ljung-Box weights do not offer an advantage under either hypothesis.

[2] The penalty $q_k = \sqrt{k-1}$ also satisfies a required lower bound on q_k arising from a probability bound used to tackle a maximum operator over an unbounded asymptotic set of (h,k): see Jin, Wang and Wang (2015, eq. (3.4)).

[3] See, for example, class $\mathfrak{K}_2$ in Andrews (1991), or class $\mathfrak{K}$ in de Jong and Davidson (2000, Assumption 1).

Finally, a power improvement may be yielded by combining bases, in particular for bases with uniquely defined systematic samples. Let $\mathcal{M}_T(\mathcal{B}_j)$ be max-statistics based on $\mathcal{J} \in \mathbb{N}$ orthonormal bases $\mathcal{B}_{j,k}(x)$, $j = 1, ..., \mathcal{J}$. Then ignoring penalties and weights (to ease notation here), define the so-called "*max-max-statistic*":

$$\check{\mathcal{M}}_T \equiv \max_{1 \leq j \leq \mathcal{J}} \left\{ \mathcal{M}_T(\mathcal{B}_j) \right\}. \tag{2.6}$$

As an example, we study $\check{\mathcal{M}}_T \equiv \max \{\mathcal{M}_T(\mathcal{W}), \mathcal{M}_T(\Psi)\}$ in simulation work, where $\mathcal{M}_T(\mathcal{W})$ and $\mathcal{M}_T(\Psi)$ use Walsh and composite Haar bases respectively. An asymptotic theory for $\check{\mathcal{M}}_T$ and its bootstrapped p-value follow directly from results given below since $\mathcal{J}$ is a finite constant. Other options for basis combinations are clearly available. Consider discretized bases $B_{j,k}(t)$ and the set $\{\bar{B}_{\bar{k}}(t)\}_{\bar{k}=1}^{\bar{\mathcal{K}}} = \{B_{j,k}(t) : j \in \mathcal{J}^*; k \in \mathcal{K}^*\}$ where $\mathcal{J}^*$ and $\mathcal{K}^*$ are index subsets of $\{1, ..., \mathcal{J}\}$ and $\{1, ..., \mathcal{K}\}$ yielding unique $\mathcal{B}_{j,k}(x)$ $\forall x$. Test statistics can then be derived from $\{\bar{B}_{\bar{k}}(t)\}_{\bar{k}=1}^{\bar{\mathcal{K}}}$.

3. Asymptotic theory

Write

$$z_t(h,k) \equiv \{X_t X_{t+h} B_k(t) - E[X_t X_{t+h}] B_k(t)\} \tag{3.1}$$

$$\mathcal{Z}_T(h,k) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} z_t(h,k), \tag{3.2}$$

and define a variance function

$$\sigma_T^2(h,k) \equiv E\left[\mathcal{Z}_T(h,k)^2\right].$$

In the general case $E[X_t] \in \mathbb{R}$ replace X_t with $X_t - E[X_t]$.

The main result of this section delivers a class of sequences $\{\mathcal{H}_T, \mathcal{K}_T\}$, and an array of random variables $\{\mathbf{Z}_T(h,k) : T \in \mathbb{N}\}_{h \geq 0, k \geq 1}$ normally distributed $\mathbf{Z}_T(h,k) \sim N(0, \sigma_T^2(h,k))$, such that the Kolmogorov distance

$$\rho_T \equiv \sup_{z \geq 0} \left| P\left(\max_{\mathcal{H}_T, \mathcal{K}_T} |\mathcal{Z}_T(h,k)| \leq z\right) - P\left(\max_{\mathcal{H}_T, \mathcal{K}_T} |\mathbf{Z}_T(h,k)| \leq z\right)\right| \to 0. \tag{3.3}$$

The approximation does not require standardized $\mathcal{Z}_T$ and $\mathbf{Z}_T$ in view of non-degeneracy Assumption 1.c below. We then apply the approximation to the max-correlation difference statistic.

Define σ-fields

$$\mathcal{F}_{T,t}^{\infty} \equiv \sigma\left(\{X_t X_{t+h} : 0 \leq h \leq \mathcal{H}_T\}_{\tau \geq t}\right) \text{ and } \mathcal{F}_{T,-\infty}^{t} \equiv \sigma\left(\{X_t X_{t+h} : 0 \leq h \leq \mathcal{H}_T\}_{\tau \leq t}\right),$$

and α-mixing coefficients (Rosenblatt, 1956), $\alpha_l \equiv \limsup_{T \to \infty} \sup_{t \in \mathbb{Z}} \sup_{\mathcal{A} \subset \mathcal{F}_{T,-\infty}^t, \mathcal{B} \subset \mathcal{F}_{T,t+l}^{\infty}} |\mathcal{P}(\mathcal{A} \cap \mathcal{B}) - \mathcal{P}(\mathcal{A})\mathcal{P}(\mathcal{B})|$, for $l > 0$.

We work in the setting of Chang, Chen and Wu (2021) who deliver high dimensional central limit theorems for possibly non-stationary mixing sequences or under a physical dependence setting similar to Zhang and Wu (2017). Chernozhukov, Chetverikov and Kato (2013, 2015, 2017) significantly improve on results in the literature on maxima of a high dimensional sample mean of stationary independent data. Chernozhukov, Chetverikov and Kato (2014, Appendix B), cf. Chernozhukov, Chetverikov and Kato (2019, Supplemental Appendix), allow for *almost surely* bounded stationary β-mixing data.

Zhang and Wu (2017) extend results in Chernozhukov, Chetverikov and Kato (2013) to a large class of dependent stationary processes. Stationarity is not suitable here since even under the null we want to allow for global non-stationarity.

Assumption 1.

*a. (*geometric mixing*): $\{X_t\}$ is α-mixing with coefficients $\alpha_l = O(\exp\{-l^{\phi}\})$ for some $\phi > 0$.*

*b. (*subexponential tails*): $\max_{1\leq t\leq T} P(|X_t| > c) \leq \varpi \exp\{-c^{\vartheta_1}\mathcal{E}_T^{-\vartheta_2}\}$ for some $\varpi \geq 1$, $\vartheta_1 \geq 2\vartheta_2$ and $\vartheta_2 \geq 1$, and some sequence of constants $\{\mathcal{E}_T\}$, $\liminf_{T\to\infty} \mathcal{E}_T \geq 1$.*

*c. (*nondegeneracy*): $\liminf_{T\to\infty} E[\mathcal{Z}_T^2(h,k)] > 0\ \forall(h,k)$.*

*d. (*orthonormal basis*): $\{\mathcal{B}_k(x) : 0 \leq k \leq \mathcal{K}\}$ forms a complete orthonormal basis on $\mathcal{L}[0,1)$; $\mathcal{B}_k(x) \in \{-1,1\}$ on $[0,1)$; and $|\sum_{t=1}^{T} B_k(t)| = O(\eta(k))$ for some positive strictly monotonic function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$, $\eta(k) \nearrow \infty$ as $k \to \infty$.*

Remark 1. A version of (a)-(c) are imposed in Chang, Chen and Wu (2021, Conditions 1-3) for their Theorem 1. Their Condition 1 implies $\max_{1\leq t\leq T} E[\exp\{|X_t|^{\vartheta}\mathcal{B}_T^{-\vartheta}\}] \leq 2$ for some $\vartheta \geq 1$ and $\mathcal{E}_T \geq 1$, hence from Markov's inequality

$$\max_{1\leq t\leq T} P(|X_t| > c) \leq 2\exp\{-c^{\vartheta}\mathcal{E}_T^{-\vartheta}\}. \tag{3.4}$$

(b) generalizes their Condition 1 to ensure r-tuples $\max_{1\leq t_1,\ldots,t_r\leq T} P(|X_{t_1}\cdots X_{t_r}| > c) \leq r\varpi\exp\{-c^{\vartheta_2}\mathcal{E}_T^{-\vartheta_2}\}$, cf. Lemma A.2 in Appendix A. This is required for higher order asymptotics for the bootstrapped p-value. The specific $\varpi = 2$ in Chang, Chen and Wu (2021) is cosmetic and assured here by (a) and Lemma A.2. Further, (b) implies $\max_{1\leq t\leq T} E|X_t|^r = O(1)\ \forall r \geq 1$ by standard arguments, ruling out monotonic trend in higher moments, but clearly permitting general forms of global nonstationarity.

Remark 2. Nondegeneracy (c) is common in the time series literature (e.g. Doukhan, 1994, Theorem 1), in particular for non-standardized statistics involving nonstationary sequences.[4] In our high dimensional setting, (c) is required for Theorem 1 in Chang, Chen and Wu (2021), due to their use of Nazarov's (2003) inequality (c.f. Chernozhukov, Chetverikov and Kato, 2017, Lemma A.1). It classically rules out degenerate dispersion and deviant negative co-dependence within the sequence $\{X_tX_{t+h} - E[X_tX_{t+h}]\}_{t=1}^{T-h}$. Simply note $E[\mathcal{Z}_T^2(h,k)] = ((T-h)/T) \times E[(\boldsymbol{\lambda}_{T-h}'\check{\boldsymbol{X}}_{T-h})^2]$ where $\check{\boldsymbol{X}}_{T-h} \equiv [X_1X_{1+h} - E[X_1X_{1+h}]]_{t=1}^{T-h}$ and $\boldsymbol{\lambda}_{T-h} \equiv (T-h)^{-1/2}[B_k(1),\ldots,B_k(T-h)]$. Notice $\boldsymbol{\lambda}_{T-h}'\boldsymbol{\lambda}_{T-h} = 1$ since $B_k^2(t) = 1$. Thus (c) is satisfied by a classic positive definiteness property: $\inf_{\lambda'\lambda=1} E[(\lambda'\check{\boldsymbol{X}}_{T-h})^2] > 0\ \forall(h,k)$ and $\forall T \geq \underline{T}$ and some $\underline{T} \in \mathbb{N}$. For example, impose fourth order stationarity (and therefore the null), and white noise $E[X_tX_{t+h}] = 0\ \forall h \geq 1$ to reduce notation. Now define fourth order correlation coefficients $r(a,b,c,d) \equiv E[X_aX_bX_cX_d]/E[X_a^2X_c^2]$. Then by expanding $E[(\lambda'\check{\boldsymbol{X}}_{T-h})^2]$, (c) holds under pointwise non-degeneracy $E[X_1^2X_{1+h}^2] > 0$, and $\inf_{\lambda'\lambda=1}\{1 + \sum_{i=1}^{T-h-1} r(0,h,i,i+h)\sum_{t=1}^{T-h-i}\lambda_t\lambda_{t+i}\} > 0\ \forall(h,k)$, $\forall T \geq \underline{T}$, ruling out deviant negative linear dependence. See also the discussion in Chang, Chen and Wu (2021, p. 4-5). We cannot, however, impose fourth order stationarity *broadly*, and thus the preceding sufficient conditions, because that rules out an asymptotic analysis

[4]See, e.g., de Jong (1997, Assumption 2.a), but also see Billingsley (1999, Theorem 19.1).

under local or global alternatives, cf. Section 3.1. We need Assumption 1 to cover each hypothesis and therefore all asymptotic theory.

Remark 3. *If X_t is (locally) sub-Gaussian then $\mathcal{E}_T = O(1)$ and $\vartheta_1 = 2$, and under sub-exponentiality $\mathcal{E}_T = O(1)$ and $\vartheta_1 = 1$.*[5] *Sub-exponentiality is equivalent to the existence of a moment generating function (in a neighborhood of zero), hence the existence of all moments (e.g. Vershynin, 2018, Proposition 2.7.).*

Remark 4. As discussed above, the high dimensional limit theory and Gaussian approximation literatures typically assume global stationarity which would be a severe hindrance here. Even in the broad literature there are trade-offs, akin to the implied exponential moment bound in (b). In Chernozhukov, Chetverikov and Kato (2014, Appendix B), for example, X_t can be stationary β-mixing, provided $\max_{1 \leq t \leq T} \max_{\mathcal{H}_T} |X_t X_{t-h} - \gamma_h| \leq \mathcal{D}_T$ where $\mathcal{D}_T \to \infty$ ultimately restricts the maximum lag rate $\mathcal{H}_T \to \infty$. Zhang and Wu (2017) allow for unbounded functionally dependent and stationary $\{X_t X_{t+h}\}$ as long as $X_t X_{t+h}$ is a measurable function of iid random variables, and a set of technical conditions restricting dependence in high dimension hold (see their Theorem 3.2). Notice Chang, Chen and Wu's (2021) bound $\mathcal{E}_T$ in (3.4) provides a significant improvement over Chernozhukov, Chetverikov and Katos' (2014) upper bound $\mathcal{D}_T$ since clearly (3.4) allows for unbounded sequences.

Remark 5. Assumption 1 reveals a trade-off vis-à-vis JWW. We allow for nonlinear processes $\{X_t\}$ with possibly non-iid errors, and possibly global nonstationarity under the null, but X_t must have a moment generating function and exhibit geometric dependence. The latter rules out conventional GARCH processes (which lack higher moments), but includes GARCH-type processes with errors that have bounded support. JWW focus exclusively on linear processes $X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$ with iid Z_t where $E|Z_t|^{4\upsilon} < \infty$ for some $\upsilon > 1$, excluding important nonlinear and conditionally heteroscedastic processes. They impose $\psi_i = O(1/[i(\ln i)^{1+\kappa}])$ for some $\kappa > 0$ and strict stationarity under the null, yielding $\sum_{h=1}^{\infty} |\gamma_h| < \infty$ since $\gamma_h = o(1/h)$. Thus JWW allow for hyperbolic *and* geometric memory decay and the possible nonexistence of higher moments.

Remark 6. The bound $|\sum_{t=1}^{T} B_k(t)| = O(\eta(k))$ in (d) is generally driven by the number of zero crossings on $[0, 1)$ in the underlying smooth basis function $\mathcal{B}_k(x)$. Indeed, by Lemma 3 in JWW, Walsh $\mathcal{W}_k(t)$ exhibit up to k zero crossings, and $|\sum_{t=1}^{T} \mathcal{W}_k(t)| \leq k + 1$ hence $\eta(k) = k$. Conversely it is easily seen that Haar composite $\Psi_k(t)$ exhibits up to 2^k zero crossings, and $|\sum_{t=1}^{T} \Psi_k(t)| = O(2^k)$ by Lemma A.1, hence $\eta(k) = 2^k$.

Recall ϕ appears in the mixing rate $\alpha_l = O(\exp\{-l^{\phi}\})$, cf. Assumption 1.a, and recall $\mathcal{E}_T$ in the Assumption 1.b subexponential tail structure.

Lemma 3.1. *Under Assumption 1, $\rho_T \leq T^{-1/9}\{\mathcal{E}_T^{2/3}[\ln(\mathcal{H}_T \mathcal{K}_T)]^{(1+2\phi)/(3\phi)} + \mathcal{E}_T[\ln \mathcal{H}_T \mathcal{K}_T]^{7/6}\} \to 0$, for any sequences $\{\mathcal{E}_T, \mathcal{H}_T, \mathcal{K}_T\}$ with $0 \leq \mathcal{H}_T \leq T - 1$, $\mathcal{H}_T = o(T)$, $\mathcal{K}_T = o(T^{\kappa})$ for some finite $\kappa > 0$, $\eta(\mathcal{K}_T) = o(\sqrt{T})$ where $\eta(\cdot)$ is the Assumption 1.d discrete basis summand bound, and*

$$\mathcal{E}_T = o\left(T^{1/6}/\{\ln(T)\}^{(1+2\phi)/(2\phi)}\right). \tag{3.5}$$

[5] Recall z is sub-Gaussian when $P(|z| > c) \leq \mathcal{K}\exp\{-\vartheta c^2\}$ for some $\vartheta, \mathcal{K} > 0$, and sub-exponential when $P(|z| > c) \leq \mathcal{K}\exp\{-\vartheta c\}$. *Local* sub-Gaussianicity allows for a non-zero mean and imposes an upper bound for only *some* c (Chareka, Chareka and Kennedy, 2006).

In this case $\max_{\mathcal{H}_T,\mathcal{K}_T} |\mathcal{Z}_T(h,k)| \xrightarrow{d} \max_{h,k\in\mathbb{N}} |\mathbf{Z}(h,k)|$ *where* $\mathbf{Z}(h,k) \sim N(0, \lim_{T\to\infty} \sigma_T^2(h,k))$ *and* $\lim_{T\to\infty} \sigma_T^2(h,k) < \infty$.

Remark 7. In a time series setting $\mathcal{H}_T = o(T)$ must hold to ensure consistency of sample autocovariances (and therefore consistency of the proposed test). We require the orthonormal basis $\mathcal{B}(x)$ bound function $\eta(\cdot)$ and maximum systematic sample counter $\mathcal{K}_T$ to satisfy $\eta(\mathcal{K}_T) = o(\sqrt{T})$ to ensure the mean summation $\mathcal{S}_T^{(k)}(h) \equiv 1/\sqrt{T}\sum_{t=1}^{T-h} E[X_t X_{t+h}]B_k(t)$ is negligible in the proof of Theorem 3.3 below. Simply note that under H_0 and Assumption 1.d, $|\mathcal{S}_T^{(k)}(h)| \leq \gamma_h |1/\sqrt{T}\sum_{t=1}^{T-h} B_k(t)| \leq \gamma_h \eta(k)/\sqrt{T}$ by Assumption 1.d. Thus $\max_{1\leq k\leq \mathcal{K}_T} |\mathcal{S}_T^{(k)}| \leq \gamma_h \eta(\mathcal{K}_T)/\sqrt{T} \to 0$ when $\eta(\mathcal{K}_T) = o(\sqrt{T})$. We also exploit $\mathcal{H}_T = o(T)$ and the more concrete $\mathcal{K}_T = o(T^\kappa)$ for some finite $\kappa > 0$ (that may be arbitrarily large) to ensure a high dimensional central limit theorem applies. The latter $\mathcal{K}_T = o(T^\kappa)$ is implied by $\eta(\mathcal{K}_T) = o(\sqrt{T})$ for Walsh and Haar functions: see below. Together with (3.5) this yields the Kolmogorov distance $\rho_T \to 0$. Theory developed in Chang, Chen and Wu (2021, Theorem 1), however, allows for a significantly greater (exponential) upper bound on the product $\mathcal{H}_T\mathcal{K}_T$ for general high dimensional means with dimension $\mathcal{H}_T\mathcal{K}_T$.

Remark 8. Walsh functions $\mathcal{W}_k(t)$ have $\eta(k) = k$ hence $\mathcal{K}_T = o(\sqrt{T})$, while Haar composite $\Psi_k(t)$ have $\eta(k) = 2^k$ hence $\mathcal{K}_T = o(\ln(T))$, yielding $\mathcal{K}_T = o(T^\kappa)$ respectively for some, or any, $\kappa > 0$.

Remark 9. The result reveals a memory/heterogeneity trade-off: as $\phi \searrow 0$ such that geometric mixing memory deepens, the maximum allowed rate $\mathcal{E}_T \to \infty$ is slower, restricting the range of feasible exponential tails.

The following corollary focuses on the case $\mathcal{E}_T = O(1)$ which automatically satisfies (3.5) e.g. when X_t is sub-exponential.

Corollary 3.2. *Let Assumption 1 hold with* $\mathcal{E}_T = O(1)$. *Then* $\rho_T \to 0$ *for any sequences* $\{\mathcal{H}_T, \mathcal{K}_T\}$ *with* $\mathcal{H}_T = o(T)$, $\mathcal{K}_T = o(T^\kappa)$ *for some finite* $\kappa > 0$, *and* $\eta(\mathcal{K}_T) = o(\sqrt{T})$.

Now define

$$\sigma^2(h,k) \equiv \lim_{T\to\infty} \sigma_T^2(h,k).$$

Under H_0 and Assumption 1, $\sigma^2(h,k) \in (0,\infty)$. We now have a limit theory for the max-correlation difference.

Theorem 3.3. *Let* H_0 *and Assumption 1 hold, and let* $\mathcal{H}_T, \mathcal{K}_T \to \infty$. *Let* $\{\mathbf{Z}(h,k) : h,k \in \mathbb{N}\}$ *be a zero mean Gaussian process with* $\mathbf{Z}(h,k) \sim N(0,\sigma^2(h,k))$. *Then it holds that* $\mathcal{M}_T \xrightarrow{d} \gamma_0^{-1} \max_{h,k\in\mathbb{N}} |\mathbf{Z}(h,k)|$ *for any* $\{\mathcal{E}_T, \mathcal{H}_T, \mathcal{K}_T\}$ *with* $\mathcal{H}_T = o(T)$, $\mathcal{K}_T = o(T^\kappa)$ *for some finite* $\kappa > 0$, $\eta(\mathcal{K}_T) = o(\sqrt{T})$ *and (3.5).*

Remark 10. Consider the weighted/penalized version $\mathcal{M}_T^{(w,p)}$ in (2.5), and assume the weights satisfy $\liminf_{T\to\infty} \inf_{\mathcal{H}_T,\mathcal{K}_T} \mathfrak{W}_{T,h}^{(k)} > 0$ *a.s.*, and $\max_{\mathcal{H}_T,\mathcal{K}_T} |\mathfrak{W}_{T,h}^{(k)} - \mathfrak{W}_h^{(k)}| \xrightarrow{p} 0$ where non-stochastic $\mathfrak{W}_h^{(k)}$ satisfy $\inf_{h,k\in\mathbb{N}} \mathfrak{W}_h^{(k)} > 0$. The penalty functions (p_v, q_v) are positive, monotonically increasing and bounded on compact sets. Then from arguments used to prove Theorem 3.3, it follows that under the conditions of Theorem 3.3:

$$\mathcal{M}_T^{(w,p)} \xrightarrow{d} \gamma_0^{-1} \max_{k \in \mathbb{N}} \left[\max_{h \in \mathbb{N}} \left\{ \mathfrak{W}_h^{(k)} \left| \mathbf{Z}(h,k) \right| - p_h \right\} - q_k \right]$$

Now suppose we standardize with $\mathfrak{W}_{T,h}^{(k)} = 1/\hat{\mathcal{V}}_T(h,k)$ with HAC estimator $\hat{\mathcal{V}}_T^2(h,k)$ in (2.4), and kernel function $\mathcal{K}(\cdot)$ belonging to class $\mathfrak{K}$ in de Jong and Davidson (2000, Assumption 1), or class $\mathfrak{K}_2 \supset \mathfrak{K}$ in Andrews (1991). de Jong and Davidson (2000) allow for possibly globally nonstationary mixing sequences (or non-mixing satisfying a *near epoch dependence* property). In their environment with bandwidth $\beta_T = o(T)$ we have $\mathfrak{W}_{T,h}^{(k)} > 0 \; a.s.$ and $\mathfrak{W}_{T,h}^{(k)} \xrightarrow{p} \mathfrak{W}_h^{(k)} = 1/\mathcal{V}(h,k)$ where $\mathcal{V}^2(h,k) = \gamma_0^{-1} \lim_{T\to\infty} \sigma_T^2(h,k)$. Uniformity $\max_{\mathcal{H}_T,\mathcal{K}_T} |\mathfrak{W}_{T,h}^{(k)} - \mathfrak{W}_h^{(k)}| \xrightarrow{p} 0$ can be proved using theory developed in this section, and Section 4, omitted here for space considerations.

3.1. Max-correlation difference under H_1

The correlation difference expands to:

$$\sqrt{T}(\hat{\rho}_h^{(k)} - \hat{\rho}_h) = \frac{1}{\hat{\gamma}_0} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \left(X_t X_{t+h} - E\left[X_t X_{t+h} \right] \right) B_k(t) + \frac{1}{\hat{\gamma}_0} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} E\left[X_t X_{t+h} \right] B_k(t). \tag{3.6}$$

Under either hypothesis $1/\sqrt{T} \sum_{t=1}^{T-h} (X_t X_{t+h} - E[X_t X_{t+h}]) B_k(t)$ is asymptotically normal. For the sample variance, we similarly have under either hypothesis and Assumption 1:

$$\sqrt{T} \left(\hat{\gamma}_0 - \frac{1}{T} \sum_{t=1}^{T} E\left[X_t^2 \right] \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(X_t^2 - E\left[X_t^2 \right] \right) = O_p(1).$$

Hence, $\hat{\gamma}_0 = g_0 + O_p(1/\sqrt{T})$ assuming existence of $g_0 \equiv \lim_{T\to\infty} 1/T \sum_{t=1}^{T} E[[X_t^2]$. See below for derivations of g_0 under local and global alternatives.

In order to handle $1/\sqrt{T} \sum_{t=1}^{T-h} E[X_t X_{t+h}] B_k(t)$ in (3.6), we need a representation of a non-stationary covariance for fixed and local alternatives. Let $\gamma_h(u)$ be the time varying autocovariance function on $[0,1]$. In the framework of locally stationary processes (cf. Dahlhaus, 1997, 2009), we may state the global alternative hypothesis as

$$H_1 : \int_0^1 \left(\gamma_h(u) - \int_0^1 \gamma_h(v) dv \right)^2 du > 0 \text{ for some } h \geq 0. \tag{3.7}$$

Thus under H_1 there exists a lag h and subset $\mathcal{S}_h \subset [0,1]$ with positive Lebesgue measure such that $\gamma_h(u) \neq \int_0^1 \gamma_h(v) dv$ on $\mathcal{S}_h$; hence $\gamma_h(u)$ is not *almost everywhere* constant on $[0,1]$.

Now, by completeness of $\{\mathcal{B}_k(u) : 0 \leq k \leq \mathcal{K}\}$ under Assumption 1.d, we may write $\gamma_h(u) = \sum_{k=0}^{\infty} \omega_{h,k} \mathcal{B}_k(u) = \omega_{h,0} + \sum_{k=1}^{\infty} \omega_{h,k} \mathcal{B}_k(u)$, where $\omega_{h,k} = \int_0^1 \gamma_h(u) \mathcal{B}_k(u)$ by orthonormality. Hence, under H_1 and orthonormality, for some $h \geq 0$.

$$\int_0^1 \left(\gamma_h(u) - \int_0^1 \gamma_h(v) dv \right)^2 du = \int_0^1 \left(\sum_{k=1}^{\infty} \omega_{h,k} \mathcal{B}_k(u) \right)^2 du = \sum_{k=1}^{\infty} \omega_{h,k}^2 > 0,$$

which yields under H_1, $\max_{h,k\in\mathbb{N}} |\int_0^1 \gamma_h(u)\mathcal{B}_k(u)| > 0$.

A sequence of local alternatives with $\sqrt{T}$-drift logically follows. Let

$$H_1^L : E\left[X_t X_{t+h}\right] = \gamma_h + c_h(t/T)/\sqrt{T}, \tag{3.8}$$

where γ_h is a constant for each h, $\max_{h\in\mathbb{N}} |\gamma_h| \le K < \infty$, and $c_h : [0,1] \to \mathbb{R}$ are integrable functions on $[0,1]$ uniformly over h (i.e. $\sup_{h\in\mathbb{N}} |\int_0^1 c_h(u)du)| < \infty$), that satisfy (3.7) Thus, under local alternative (3.8), by the preceding discussion:

$$\liminf_{T\to\infty} \max_{h,k\in\mathbb{N}} \left| \int_0^1 c_h\left(u\right)\mathcal{B}_k(u)du \right| > 0. \tag{3.9}$$

In order to ensure $\min_{t\in\mathbb{Z}} E[X_t^2] > 0$, assume $\gamma_0 > 0$ and $c_0(u) \ge 0$ *almost everywhere*. Notice $\lim_{T\to\infty} |T^{-1}\sum_{t=1}^T c_0(t/T)| = |\int_0^1 c_0(u)du| < \infty$ yields under H_1^L:

$$g_0 \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T E[X_t^2] = \gamma_0 + \lim_{T\to\infty} \frac{1}{\sqrt{T}}\frac{1}{T}\sum_{t=1}^T c_0(t/T) = \gamma_0.$$

Under Assumption 1.d $|\sum_{t=1}^T B_k(t)| = O(\eta(k))$, and $\max_{h\in\mathbb{N}} |\gamma_h| \le K$ and $\eta(\mathcal{K}_T) = o(\sqrt{T})$ by supposition. Hence $\max_{\mathcal{H}_T,\mathcal{K}_T} |1/\sqrt{T}\sum_{t=1}^{T-h} B_k(t)| = o(1)$. Thus under H_1^L:

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} E\left[X_t X_{t+h}\right] B_k(t) &= \gamma_h \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} B_k(t) + \frac{1}{T}\sum_{t=1}^{T-h} c_h(t/T)B_k(t) \\ &= o\left(1\right) + \frac{1}{T}\sum_{t=1}^{T-h} c_h(t/T)B_k(t) \to \int_0^1 c_h\left(u\right)\mathcal{B}_k(u)du, \end{aligned} \tag{3.10}$$

where here and below $o(1)$, and all subsequent $O_p(\cdot)$ and $o_p(\cdot)$ terms, do not depend on (h,k).

Asymptotics in our mixing setting rest on uniform limit theory over (h,k), which here needs to extend to the limit in (3.10). We therefore enhance local alternative (3.8) by assuming $c_h(\cdot)$ satisfies for any $\{\mathcal{H}_T, \mathcal{K}_T\}$:

$$\max_{\mathcal{H}_T,\mathcal{K}_T} \left| \frac{1}{T}\sum_{t=1}^{T-h} c_h(t/T)B_k(t) - \int_0^1 c_h\left(u\right)\mathcal{B}_k(u)du \right| \to 0. \tag{3.11}$$

Now define:

$$C(h,k) = \int_0^1 c_h\left(u\right)\mathcal{B}_k(u)du. \tag{3.12}$$

Then under H_1^L, $\liminf_{T\to\infty} \max_{h,k\in\mathbb{N}} |C(h,k)| > 0$ in view of (3.9). Use arguments in the proof of Theorem 3.3 to yield under H_1^L:

$$\begin{aligned} \sqrt{T}\left(\hat{\rho}_h^{(k)} - \hat{\rho}_h\right) &= \frac{1}{g_0}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left(X_t X_{t+h} - E\left[X_t X_{t+h}\right]\right) B_k(t) \\ &+ \frac{1}{g_0}\left(\int_0^1 c_h\left(u\right)\mathcal{B}_k(u)du + o\left(1\right)\right) + O_p(1/\sqrt{T}), \end{aligned}$$

hence by Lemma 3.1, for any $\{\mathcal{E}_T, \mathcal{H}_T, \mathcal{K}_T\}$ with $\mathcal{H}_T = o(T)$, $\mathcal{K}_T = o(T^\kappa)$ for some finite $\kappa > 0$, $\eta(\mathcal{K}_T) = o(\sqrt{T})$ and (3.5):

$$\max_{\mathcal{H}_T, \mathcal{K}_T} \left| \sqrt{T} \left(\hat{\rho}_h^{(k)} - \hat{\rho}_h \right) \right| \xrightarrow{d} \frac{1}{g_0} \max_{h,k \in \mathbb{N}} |\mathbf{Z}(h,k) + C(h,k)| \tag{3.13}$$

Thus the max-correlation difference test has non-negligible power under the sequence of $\sqrt{T}$-local alternatives (3.8) when $c_h(\cdot)$ satisfy (3.7), for any complete orthonormal basis in view of (3.9). Notice under H_0 we have $g_0 = \gamma_0$, and $c_h(u) = 0\ \forall u, h$ so that $C(h,k) = 0\ \forall h, k$, yielding Theorem 3.3.

As a global generalization of H_1^L, we may write H_1 in discrete form as

$$H_1 : E\left[X_t X_{t+h}\right] = \gamma_h + c_h(t/T), \tag{3.14}$$

where as above $c_h(\cdot)$ satisfies (3.7). In this case $g_0 \equiv \lim_{T\to\infty} T^{-1} \sum_{t=1}^T E[X_t^2]$ is identically:

$$g_0 = \gamma_0 + \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} c_0(t/T) = \gamma_0 + \int_0^1 c_0(u)\, du > 0.$$

Repeating the above derivations, we find similar to (3.10),

$$\frac{1}{T} \sum_{t=1}^{T-h} E\left[X_t X_{t+h}\right] B_k(t) = \gamma_h \frac{1}{T} \sum_{t=1}^{T-h} B_k(t) + \frac{1}{T} \sum_{t=1}^{T-h} c_h(t/T) B_k(t) = \int_0^1 c_h(u)\, \mathcal{B}_k(u) du + o(1).$$

and therefore

$$\sqrt{T}\left(\hat{\rho}_h^{(k)} - \hat{\rho}_h\right) = \frac{1}{g_0} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \left(X_t X_{t+h} - E\left[X_t X_{t+h}\right]\right) B_k(t)$$
$$+ \frac{1}{g_0} \sqrt{T} \left(\int_0^1 c_h(u)\, \mathcal{B}_k(u) du + o(1) \right) + O_p(1/\sqrt{T}).$$

Thus $\max_{\mathcal{H}_T, \mathcal{K}_T} |\sqrt{T}(\hat{\rho}_h^{(k)} - \hat{\rho}_h)| \xrightarrow{p} \infty$ given $\liminf_{T\to\infty} \max_{\mathcal{H}_T, \mathcal{K}_T} | \int_0^1 c_h(u)\, \mathcal{B}_k(u) du| > 0$.

The next result summarizes the preceding discussion.

Theorem 3.4. *Let Assumption 1 hold, and let* $\{\mathcal{E}_T, \mathcal{H}_T, \mathcal{K}_T\}$ *satisfy* $0 \leq \mathcal{H}_T \leq T - 1$, $\mathcal{H}_T \to \infty$, $\mathcal{K}_T \to \infty$, $\mathcal{H}_T = o(T)$, *and (3.5).*

a. Under H_1^L, *(3.13) holds for non-zero* $C(h,k)$ *in (3.12), and any sequence* $\{\mathcal{K}_T\}$ *with* $\mathcal{K}_T = o(T^\kappa)$ *for some finite* $\kappa > 0$ *and* $\eta(\mathcal{K}_T) = o(\sqrt{T})$.

b. Under H_1, $\max_{\mathcal{H}_T, \mathcal{K}_T} |\sqrt{T}(\hat{\rho}_h^{(k)} - \hat{\rho}_h)| \xrightarrow{p} \infty$ *for any* $\{\mathcal{K}_T\}$.

In the following example we study a simple break in variance in order to show how the max-test behaves asymptotically.

Example 3.5 (Structural Break in Variance). Assume covariances do not depend on time: $E[X_t X_{t-h}] = \gamma_h$ for every $h \geq 1$, but there is a structural break in variance at mid-sample (for simplicity of discussion), cf. Perron (2006):

$$E[X_t^2] = g_{1,T} \text{ for } t = 1, ..., [T/2] \text{ and } E[X_t^2] = g_{2,T} \text{ for } t = [T/2] + 1, ..., T$$

for some strictly positive finite sequences $\{g_{1,T}, g_{2,T}\}$, $g_{1,T} \neq g_{2,T}$. In terms of, e.g., Walsh or composite Haar systematic samples and H_1^L, this translates to

$$c_0(u) = c_{0,1} > 0 \text{ for } u \in [0, 1/2), \text{ and } c_0(u) = c_{0,2} > 0 \text{ for } u \in [1/2, 1],$$

where $c_{0,1} \neq c_{0,2}$. All other $c_h(u) = 0$ on $[0, 1]$, $h \geq 1$. Hence, by construction of the first Walsh function $W_1(u)$ (or Haar composite $\psi_1(u)$):

$$\int_0^1 c_0(u)\, W_1(u) du = \int_0^{1/2} c_0(u)\, du - \int_{1/2}^1 c_0(u)\, du = \frac{c_{0,1} - c_{0,2}}{2} \neq 0.$$

Further:

$$g_0 \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[X_t^2] = \gamma_0 + \int_0^1 c_0(u)\, du = \gamma_0 + \frac{c_{0,1} + c_{0,2}}{2}.$$

The normalized correlation difference therefore satisfies for $h \geq 1$,

$$\sqrt{T}\left(\hat{\rho}_h^{(k)} - \hat{\rho}_h\right) = \frac{1}{g_0} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \left(X_t X_{t+h} - E\left[X_t X_{t+h}\right]\right) B_k(t) + o_p(1).$$

Under H_1 at lag $h = 0$ and $k = 1$ we then have:

$$\begin{aligned}\sqrt{T}\left(\hat{\rho}_0^{(1)} - \hat{\rho}_0\right) &= \frac{1}{g_0} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(X_t^2 - E\left[X_t^2\right]\right) \mathcal{W}_1(t) + \sqrt{T}\left(\frac{c_{0,1} - c_{0,2}}{2g_0}\right) + o_p(1) \\ &= \mathcal{Z}_T + \mathcal{C}_T + o_p(1),\end{aligned}$$

say. In view of asymptotic normality of $\mathcal{Z}_T$, and $|\mathcal{C}_T| \to \infty$, the max-correlation difference test is consistent when only the variance $E[X_t^2]$ exhibits a break given $\max_{\mathcal{H}_T, \mathcal{K}_T} \sqrt{T}|\hat{\rho}_h^{(k)} - \hat{\rho}_h| \geq \sqrt{T}|\hat{\rho}_0^{(1)} - \hat{\rho}_0| = \sqrt{T}|\mathcal{Z}_T + \mathcal{C}_T| \xrightarrow{p} \infty$.

4. Dependent wild bootstrap

We exploit a blockwise wild (multiplier) bootstrap for p-value approximation (cf. Liu, 1988). The method appears in various places as a multiplier bootstrap extension of block-based bootstrap methods (e.g. Künch, 1989). Shao (2010) presents a general nonoverlapping dependent wild bootstrap, exploiting a class of kernel smoothing weights that omits the truncated kernel, and uses only "big" blocks of data ("little" block size is effectively zero). Shao (2011) uses the same method exclusively with a truncated kernel for a white noise test for a stationary process that is a measurable function of an iid sequence. In both cases a sequence $\{X_t\}_{t=1}^T$ is decomposed into $[T/b_T]$ blocks of size $1 \leq b_T < T$, $b_T \to \infty$ and $b_T = o(T)$.

Chernozhukov, Chetverikov and Kato (2019) exploit a Bernstein-like "big" and "little" block multiplier bootstrap for high dimensional sample means of stationary, dependent and bounded sequences. They apply a wild bootstrap on big blocks and effectively remove the asymptotically negligible little blocks. Zhang and Cheng (2014) expand that method for stationary processes by using two mutually independent iid sequences, one each for big and small blocks.

We expand ideas in Shao (2011) to non-stationary sequences. The use of only one set of "big" blocks and a truncated kernel eases technical arguments and notation, but a more general use of smoothing kernels and big/little blocks is readily supported by the theory presented here.

Set a block size b_T such that $1 \leq b_T < T$, $b_T/T^{\iota} \to \infty$ and $b_T/T^{1-\iota} \to 0$ for some tiny $\iota > 0$. The number of blocks is $\mathcal{N}_T = [T/b_T]$. Denote the index blocks by $\mathfrak{B}_s = \{(s-1)b_T + 1, \ldots, sb_T\}$ with $s = 1, \ldots, \mathcal{N}_T$, and $\mathfrak{B}_{\mathcal{N}_T+1} = \{\mathcal{N}_T b_T, ..., T\}$. Generate iid random numbers $\{\xi_1, \ldots, \xi_{\mathcal{N}_T}\}$ with $E[\xi_i]$ = 0, $E[\xi_i^2] = 1$, and $E[\xi_i^4] < \infty$. Typically in practice ξ_i is iid $N(0,1)$, and we make that assumption here to shorten the proof of a key supporting Lemma A.4. See its proof in the supplemental material for further comments.

Define an auxiliary variable $\varphi_t = \xi_s$ if $t \in \mathfrak{B}_s$, and let $\Delta\hat{g}_T^{(dw)}(h,k)$ be a (centered) bootstrapped version of $\hat{\gamma}_h^{(k)} - \hat{\gamma}_h = T^{-1}\sum_{t=1}^{T-h} X_t X_{t+h} B_k(t)$:

$$\Delta\hat{g}_T^{(dw)}(h,k) \equiv \frac{1}{T}\sum_{t=1}^{T-h}\varphi_t\left\{X_t X_{t+h} B_k(t) - \frac{1}{T}\sum_{s=1}^{T-h} X_s X_{s+h} B_k(s)\right\}. \tag{4.1}$$

An asymptotically equivalent technique centers only on $X_t X_{t+h}$, the key stochastic term:

$$\Delta\hat{g}_T^{(dw)}(h,k) \equiv \frac{1}{T}\sum_{t=1}^{T-h}\varphi_t\left\{X_t X_{t+h} - \frac{1}{T}\sum_{s=1}^{T-h} X_s X_{s+h}\right\} B_k(t).$$

The bootstrapped test statistic is then $\mathcal{M}_T^{(dw)} \equiv \hat{\gamma}_0^{-1}\max_{\mathcal{H}_T,\mathcal{K}_T}|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)|$. Repeat M times. As a by-product of the main result below, conditional on the sample $\{X_t\}_{=1}^T$ this results in a sequence $\{\mathcal{M}_{T,i}^{(dw)}\}_{i=1}^M$ of iid draws $\mathcal{M}_{T,i}^{(dw)}$ from the limit null distribution of $\mathcal{M}_T$ as $T \to \infty$ asymptotically with probability approaching one. The approximate p-value is:

$$\hat{p}_{T,M}^{(dw)} \equiv \frac{1}{M}\sum_{i=1}^{M} I\left(\mathcal{M}_{T,i}^{(dw)} \geq \mathcal{M}_T\right).$$

The bootstrap test rejects H_0 at significance level α when $\hat{p}_{T,M}^{(dw)} < \alpha$.

The multiplier bootstrap has been studied in many contexts with intuitive insights given. Consult, e.g., Liu (1988), Shao (2010), and Shao (2011) to name a few. Centering $X_t X_{t+h} B_k(t) - 1/T\sum_{s=1}^{T-h} X_s X_{s+h} B_k(s)$ is required because we use $\{X_t X_{t+h} B_k(t)\}$ to approximate the null distribution, *whether it is true or not*, and $E[X_t X_{t+h} B_k(t)] \neq 0$ for some (h,k) under H_1. The block-wise independent zero-mean Gaussian multiplier φ_t serves the purpose that $\varphi_t\{X_t X_{t+h} B_k(t) - 1/T\sum_{s=1}^{T-h} X_s X_{s+h} B_k(s)\}$, conditioned on sample $\mathfrak{X}_T \equiv \{X_t\}_{t=1}^T$, is zero mean normally distributed; indeed, $\Delta\hat{g}_T^{(dw)}(h,k)|\mathfrak{X}_T \sim N(0,\mathcal{V}_T(h,k))$ for some $\mathcal{V}_T(h,k) > 0$ $a.s.$ The blocks are constructed such that the dispersion term $\mathcal{V}_T(h,k)$ well approximates the null limiting variance under general dependence, that is $\mathcal{V}_T(h,k) \xrightarrow{p} \sigma^2(h,k)$. Thus, in the jargon of Giné and Zinn (1990, Section 3), $\Delta\hat{g}_T^{(dw)}(h,k)|\mathfrak{X}_T \xrightarrow{d} \mathbf{Z}(h,k) \sim N(0,\sigma^2(h,k))$ *in probability*,[6] ensuring the bootstrapped process yields the null distribution, irrespective of whether H_0 holds or not.

[6] Thus, convergence is in distribution, *asymptotically with probability approaching one* for any drawn sample $\{X_t\}_{t=1}^T$.

Recall $z_t(h,k)$ in (3.1) and $\mathcal{Z}_T(h,k) \equiv 1/\sqrt{T}\sum_{t=1}^{T-h} z_t(h,k)$. Write $\ddot{g}_T(h,k) \equiv 1/(T-h)\sum_{u=1}^{T-h} E\left[X_u X_{u+h}\right] B_k(u)$ and

$$\mathfrak{X}_{T,l}(h,k) \equiv \sum_{t=(l-1)b_T+1}^{lb_T} \left\{X_t X_{t+h} B_k(t) - \ddot{g}_T(h,k)\right\},$$

and define pre-asymptotic and asymptotic long run covariance functions $s_T^2(h,k;\tilde{h},\tilde{k}) \equiv 1/T\sum_{l=1}^{(T-h\vee\tilde{h})/b_T} E[\mathfrak{X}_{T,l}(h,k)\mathfrak{X}_{T,l}(\tilde{h},\tilde{k})]$ and $s^2(h,k;\tilde{h},\tilde{k}) \equiv \lim_{T\to\infty} s_T^2(h,k;\tilde{h},\tilde{k})$.

Assumption 2.

a. (i) $\liminf_{T\to\infty} s_T^2(h,k;\tilde{h},\tilde{k}) > 0\ \forall(h,\tilde{h},k,\tilde{k})$; *and* (ii) $\max_{\mathcal{H}_T,\mathcal{K}_T} |s_T^2(h,k;\tilde{h},\tilde{k}) - s^2(h,k;\tilde{h},\tilde{k})| = O(T^{-\iota})$ *for some infinitessimal* $\iota > 0$.

b. $b_T/T^{\iota} \to \infty$ *and* $b_T = o(T^{1/2-\iota})$ *for some infinitessimal* $\iota > 0$.

Remark 11. $(a.i)$ is the fourth order block bootstrap version of Assumption 1.c, used to ensure a high dimensional central limit theory extends to a long run bootstrap variance, cf. Chernozhukov, Chetverikov and Kato (2013, Lemma 3.1). $(a.ii)$ seems unavoidable, and is required to link covariance functions for a high dimensional bootstrap theory, cf. Chernozhukov, Chetverikov and Kato (2013, Lemma 3.1) and Chernozhukov, Chetverikov and Kato (2015, Theorem 2, Proposition 1). The property is trivial under stationary geometric mixing, and otherwise restricts the degree of allowed heterogeneity.

Remark 12. (b) simplifies a bootstrap weak convergence proof, but can be weakened at the cost of added notation, e.g. $b_T/(\ln(T))^a \to \infty$ and $b_T = o(T^{1/2}/(\ln(T))^b)$ for some $a, b > 0$.

Remark 13. A general sacrifice is the reduced lag upper bound $\mathcal{H}_T = O(T^{1-\iota}/b_T)$ required for Theorem 4.1 below. In particular, in the proof of supporting Lemma A.4, we exploit a high dimensional multiplier bootstrap central limit theorem, which requires high dimensional uniform convergence of a bootstrap variance at rate $O(1/T^{\iota})$ for some tiny $\iota > 0$. In view of the block size $b_T \to \infty$ that naturally appears in the variance, and lagging in a covariance setting, we yield upper bounds in the proof similar to $b_T\frac{h}{T-h}$, hence $b_T\frac{h}{T-h} = O(1/T^{\iota})$ uniformly over h when $b_T\frac{\mathcal{H}_T}{T} = O(1/T^{\iota})$. This suggests a logical trade-off: we can enforce a slow block size growth rate to reach $\mathcal{H}_T/\sqrt{T} \to \infty$, although we cannot achieve $\mathcal{H}_T = o(T)$. In simulation work we find block sizes $b_T \propto T^{1/2-\iota}$ work well, effectively restricting $\mathcal{H}_T = O(T^{1/2-\iota})$.

The blockwise wild bootstrap is valid asymptotically.

Theorem 4.1. *Let Assumptions 1-2 hold, let* $\mathcal{H}_T, \mathcal{K}_T \to \infty$, *and let the number of bootstrap samples* $M = M_T \to \infty$ *as* $T \to \infty$. *Let* $\{\mathcal{E}_T, \mathcal{H}_T\}$ *satisfy* $0 \leq \mathcal{H}_T \leq T-1$, $\mathcal{H}_T = O(T^{1-\iota}/b_T)$ *and (3.5), where block size* $b_T \to \infty$ *and* $b_T = O(T^{1/2-\iota})$. *Under* H_0, $P(\hat{p}_{T,M}^{(dw)} < \alpha) \to \alpha$ *for any sequence* $\{\mathcal{K}_T\}$ *satisfying* $\mathcal{K}_T = o(T^{\kappa})$ *for some finite* $\kappa > 0$ *and* $\eta(\mathcal{K}_T) = o(\sqrt{T})$. *Under* H_1 *in (3.14) where* $c_h(\cdot)$ *satisfy (3.9),* $P(\hat{p}_{T,M}^{(dw)} < \alpha) \to 1$ *for any* $\{\mathcal{K}_T\}$.

5. Monte carlo study

We now study the proposed bootstrap test in a controlled environment. We generate 1000 independently drawn samples from various models, with sample sizes $T \in \{64, 128, 256, 512\}$. The models under the null and alternative hypotheses are detailed below.

5.1. Empirical size

We use four models of covariance stationary processes: MA(1), AR(1), and Self Exciting Threshold AR(1) [SETAR], in each case with iid or GARCH innovations; and GARCH (1,1) with iid errors.

null-1	MA(1)	$X_t = \epsilon_t$
null-2	AR(1)	$X_t = .5X_{t-1} + \epsilon_t$
null-3	SETAR	$X_t = .7X_{t-1} - 1.4X_{t-1}I(X_{t-1} > 0) + \epsilon_t,$
null-4	GARCH(1,1)	$X_t = \sigma_t z_t,\, z_t \overset{iid}{\sim} N(0,1),\ \sigma_t^2 = 1 + .3\epsilon_{t-1}^2 + .6\sigma_{t-1}^2$

Models #1-#3 have an iid error ϵ_t distributed $N(0,1)$ or Student's-t with 5 degrees of freedom (t_5); or ϵ_t is stationary GARCH(1,1) $\epsilon_t = \sigma_t z_t$, $z_t \overset{iid}{\sim} N(0,1)$, $\sigma_t^2 = 1 + .3\epsilon_{t-1}^2 + .6\sigma_{t-1}^2$, with iteration $\sigma_1^2 = 1$ and $\sigma_t^2 = 1 + .3\epsilon_{t-1}^2 + .6\sigma_{t-1}^2$ for $t = 2, ..., T$. The SETAR model switches between AR(1) regimes with correlations .7 and $-.7$. GARCH and SETAR models, and any model with GARCH errors, do not have a linear form $X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$, with iid Z_t and non-random ψ_i, and therefore do not satisfy conditions in JWW and elsewhere. We simulate $2T$ observations for each model and retain the latter T observations for analysis. Test results in GARCH cases should be viewed with caution: max-test asymptotics have only been established under sub-exponentail tails, and JWW's test requires a linear model with an iid error.

5.2. Empirical power

We study empirical power by using models similar to those used in Paparoditis (2010b); Dette, Preuß and Vetter (2011), Preuß, Vetter and Dette (2013) and JWW, with the addition of allowing for non-iid errors and non-stationarity in variance. The models are as follows:

alt-1	(NI)	$X_t = 1.1\cos\{1.5 - \cos(4\pi t/T)\}\epsilon_{t-1} + \epsilon_t$
alt-2	(NVIII)	$X_t = .8\cos\{1.5 - \cos(4\pi t/T)\}\epsilon_{t-6} + \epsilon_t$
alt-3	(NII)	$X_t = .6 \times \sin(4\pi t/T)X_{t-1} + \epsilon_t$
alt-4	(NIII)	$X_t = \begin{cases} .5X_{t-1} + \epsilon_t & \text{for } \{1 \le t \le T/4\} \cup \{3T/4 < t \le T\} \\ -.5X_{t-1} + \epsilon_t & \text{for } T/4 < t \le 3T/4 \end{cases}$
alt-5	(NVI)	$X_t = \begin{cases} .5X_{t-1} + \epsilon_t & \text{for } 1 \le t \le T/2 \\ -.5X_{t-1} + \epsilon_t & \text{for } T/2 < t \le T \end{cases}$
alt-6	(eq. (16))	$X_t = 2\epsilon_t - \{1 + .5\cos(2\pi t/T)\}\epsilon_{t-1}$
alt-7	(NV)	$X_t = -.9\sqrt{(t/T)}X_{t-1} + \epsilon_t$
alt-8		$X_t = .5X_{t-1} + v_t \colon \begin{cases} v_t = \epsilon_t & \text{for } 1 \le t \le 3T/4 \\ v_t = 2\epsilon_t & \text{for } 3T/4 < t \le T \end{cases}$
alt-9		$X_t = .8\cos\{1.5 - \cos(4\pi t/T)\}\epsilon_{t-25} + \epsilon_t$

Models 1-7 are used in JWW: we display parenthetically their corresponding model/equation number. Models 1, 2, 4 are considered in Paparoditis (2010b); Dette, Preuß and Vetter (2011) use models 1, 2, 4, and 6; and Preuß, Vetter and Dette (2013) study 2, 5, and 7. Alt-8 presents a structural change in variance only, and alt-9 is a distant version of alt-2 and therefore more difficult to detect (lag 25 as opposed to lag 6). As above, we use either iid standard normal, iid t_5, or GARCH(1,1) ϵ_t.

Table 1. $\mathcal{H}_T, \mathcal{K}_T$ Combinations by Basis

	Walsh Basis $\{W_k(t)\}$				Haar Basis $\{\Psi_k(t)\}$			
	Case 1 (JWW)		Case 2		Case 1 (JWW)		Case 2	
T	H_T	K_T	H_T	K_T	H_T	K_T	H_T	K_T
	$\log_2(T)^{.99} - 3$	$T^{1/3}$	$2T^{.49}$	$.5T^{.49}$	$\log_2(T)^{.99} - 3$	$(\ln(T))^{.99}$	$2T^{.49}$	$(\ln(T))^{.99}$
64	2	4	14	3	2	4	14	4
128	3	5	20	5	3	5	20	5
256	4	6	30	7	4	5	30	5
512	5	8	42	10	5	6	42	6

5.3. Tests

5.3.1. Max-test

We perform the bootstrapped max-correlation difference test with $\mathcal{M}_T$ and $\mathcal{M}_T^{(p)}$. The latter has penalties $p_h = (h+1)^{1/4}/2$ and $q_k = k^{1/4}/2$. More severe penalties, e.g. $q_k = k^{1/2}/2$, do not improve test performance. A weighted version of the test with HAC estimator (2.4) leads to competitive size but generally lower power, hence we focus only on $\mathcal{M}_T$ and $\mathcal{M}_T^{(p)}$. We use Walsh or Haar functions for two max-tests, and a third combined max-max-statistic shown below (2.6). We only report results based on Walsh functions because (i) the Haar-based tests (max-test, and JWW's test detailed below) yielded far lower power across most alternatives under study here; hence (ii) the max-max test performed essentially on par with the Walsh-based test.

We use 500 bootstrap samples with multiplier iid variable $\xi_t \sim N(0,1)$. Theorem 4.1 requires a block size bound $b_T = o(T^{1/2-\iota})$ for some tiny $\iota > 0$, hence we use $b_T = [T^{1/2-\eta}]$ where $\eta = 10^{-10}$. Similar block sizes, e.g. $b_T = [bT^{1/2-\eta}]$ with $b \in [.5, 2]$ lead to similar results.[7]

Theorem 4.1 also requires $\mathcal{H}_T = O(T^{1-\iota}/b_T)$, $\mathcal{K}_T = o(T^\kappa)$ for some $\kappa > 0$, and $\eta(\mathcal{K}_T) = o(\sqrt{T})$. In the Walsh case $\eta(k) = k$ hence $\mathcal{K}_T = o(\sqrt{T})$; in the Haar case $\eta(k) = 2^k$ hence $\mathcal{K}_T = o(\ln(T))$. In the Walsh case, we used two pairings of sequences $\{\mathcal{H}_T, \mathcal{K}_T\}$. The first $\mathcal{H}_T = [\log_2(T)^{.99} - 3]$ and $\mathcal{K}_T = [T^{1/3}]$ is used in JWW. The second $\mathcal{H}_T = [2T^{.49}]$ and $\mathcal{K}_T = [.5T^{.49}]$ satisfies our assumptions but are not valid in JWW. The latter $(\mathcal{H}_T, \mathcal{K}_T)$ are generally larger, where $\mathcal{H}_T$ is larger by an order of $\times 7$. This will lead to higher power for large T in theory, but in small samples obviously a larger h results in fewer observations for computation, and therefore a loss in sharpness in probability. In the Haar case we use either $\mathcal{H}_T$ above, and $\mathcal{K}_T = [(\ln(T))^{.99}]$. Refer to Table 1.

5.3.2. JWW test

Write $\hat{\boldsymbol{\gamma}}_h \equiv [\hat{\gamma}_1, ..., \hat{\gamma}_h]'$, $\hat{\boldsymbol{\gamma}}_h^{(k)} \equiv [\hat{\gamma}_1^{(k)}, ..., \hat{\gamma}_h^{(k)}]'$. The test statistic is:

$$\hat{\mathcal{D}}_T \equiv \max_{1 \leq k \leq \mathcal{K}_T} \left[\max_{1 \leq h \leq \mathcal{H}_T} \left\{ T \left(\hat{\boldsymbol{\gamma}}_h^{(k)} - \hat{\boldsymbol{\gamma}}_h \right)' \left(\hat{\Gamma}_h^{(k)} \right)^{-1} \left(\hat{\boldsymbol{\gamma}}_h^{(k)} - \hat{\boldsymbol{\gamma}}_h \right) - 2h \right\} - \sqrt{k-1} \right],$$

[7] Shao (2011) uses $b_T = [bT^{1/2}]$ with $b \in \{.5, 1, 2\}$, leading to qualitatively similar results. Hill and Motegi (2020) also use $b = 1$, but find qualitatively similar results for values $b \in \{.5, 1, 2\}$.

where $\hat{\Gamma}_h^{(k)}$ is an estimator of the $h \times h$ asymptotic covariance matrix of $\sqrt{T}(\hat{\gamma}_h^{(k)} - \hat{\gamma}_h)$. See Jin, Wang and Wang (2015, Sections 2.3-2.5) for details on computing $\hat{\Gamma}_h^{(k)}$ (under the assumption of linearity $X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$ with an iid Z_t).[8] We use both Walsh (as in JWW) and Haar bases, the same tuning parameters that JWW use for covariance matrix estimation, and the same $\{\mathcal{H}_T, \mathcal{K}_T\}$ described above.[9]

We perform the test both based on a simulated critical values (denoted $\hat{\mathcal{D}}_T^{cv}$), and bootstrapped p-values ($\hat{\mathcal{D}}_T^{dw}$) in order to make a direct comparison with the method developed here. We simulate critical values for each basis and each pair $(\mathcal{H}_T, \mathcal{K}_T)$ by running a separate simulation with $200,000$ independently drawn samples of size T of iid $N(0,1)$ distributed random variables X_t, and use the true excess kurtosis value 0 in the covariance estimator $\hat{\Gamma}_h^{(k)}$. The bootstrap is performed by replacing $\hat{\gamma}_h^{(k)} - \hat{\gamma}_h$ in $\hat{\mathcal{D}}_T$ with $\Delta \hat{g}_T^{(dw)}(h,k)$ from (4.1). We do not prove asymptotic validity of the bootstrapped p-value, but once uniform consistency of $\hat{\Gamma}_h^{(k)}$ is established, it follows identically from arguments given in the proof of Theorem 4.1. Indeed, the bootstrap is valid for linear and nonlinear processes with iid or non-iid innovations, and covering the nonstationary processes under H_1. The simulated critical values, however, are suitable in theory only for linear processes with iid innovations since they rely on the specific form of $\hat{\Gamma}_h^{(k)}$ used here, and a pivotal Gaussian null limit distribution, cf. Jin, Wang and Wang (2015, Sections 2.3-2.5).

5.4. Results

Tables A.3-A.6 in Appendix F of the supplemental material present rejection frequencies at $(1\%, 5\%, 10\%)$ significance levels when a Walsh basis is used.

The penalized max-test does not perform better than the non-penalized test, and generally performs worse under the alternative. Indeed, as discussed above, there is no theory driven reason for adding penalties for a max-test. In the sequel we therefore only discuss the non-penalized test.

Similarly, the bootstrapped JWW test is generally over-sized, and massively over-sized at small n under $(\mathcal{H}, \mathcal{K})$ Case 1, the only valid case in this study. We suspect the cause is the estimated variance matrix due to its many components and tuning parameters. We henceforth only discuss results based on simulated critical values.

5.4.1. Null

Both tests are comparable for MA and AR models with iid Gaussian or t_5 errors, with fairly accurate empirical size. The max-test has accurate size in many cases, and is otherwise conservative. JWW's test tends to be over-sized in the AR model with GARCH errors under both $(\mathcal{H}, \mathcal{K})$ cases, and is over-sized in the AR model with t_5 errors under Case 2 when $n \leq 128$. Recall $\mathcal{H}_T$ is much larger under Case 2, which will be a hindrance at smaller n for test statistics that simultaneously incorporate a set of autocovariances (e.g. Wald or portmanteau statistics).

[8] There is a typo in Jin, Wang and Wang (2015, Theorem 2) concerning their covariance matrix and therefore its estimator. A parameter κ_4, referred to as the kurtosis of the iid Z_t, is in fact the excess kurtosis (*kurtosis* -3). See Proposition 7.3.1 in Brockwell and Davis (1991), in particular eq. (7.3.5), cf. Jin, Wang and Wang (2015, p. 915).

[9] The bandwidth parameter λ in $[T^{\lambda}]$, the number of sample covariances that enter the asymptote covariance matrix estimator, is set to $\lambda = .4$ based on a private communication with the authors. Second, in order to compute the (excess) kurtosis of iid Z_t under linearity, similar to Jin, Wang and Wang (2015, eq. (15)) we use an estimator in Kreiss and Paparoditis (2015), with two bandwidths $b_j = c_j T^{-1/3}$ where $c_j = 1.25 \times$*crude scale estimate* (see Jin, Wang and Wang, 2015, p. 903). A private communication with one coauthor states the scale estimate used was $\hat{\gamma}(0)$, hence $c_j = 1.25 \times \hat{\gamma}(0)$.

Table 2. Test Dominance Summary

	$(\mathcal{H},\mathcal{K})$ Case 1			$(\mathcal{H},\mathcal{K})$ Case 2		
$H_1\backslash\epsilon_t$	$N(0,1)$	t_5	GARCH	$N(0,1)$	t_5	GARCH
alt-1	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$
alt-2	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$
alt-3	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$
alt-4	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$ small n	$\hat{\mathcal{D}}_T$
alt-5	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$
alt-6	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	similar	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	similar
alt-7	$\hat{\mathcal{D}}_T$ larger n	$\hat{\mathcal{D}}_T$ large n	$\hat{\mathcal{D}}_T$ large n	$\hat{\mathcal{D}}_T$ large n	$\hat{\mathcal{D}}_T$ large n	$\hat{\mathcal{D}}_T$ large n
alt-8	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$
alt-9	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$	$\hat{\mathcal{M}}_T$

Each cell dictates which test performed best (in certain cases). For example "$\hat{\mathcal{D}}_T$ small n" implies $\hat{\mathcal{D}}_T$ dominates for smaller sample sizes, and for other n the two tests are comparable. "$\hat{\mathcal{M}}_T$" implies $\hat{\mathcal{M}}_T$ dominates across sample sizes.

In the SETAR case JWW's test is largely over-sized, while the max-test is slightly under-sized with improvement under $(\mathcal{H},\mathcal{K})$ Case 2. JWW's test is over-sized for small n with the GARCH model, but otherwise works well.

5.4.2. Alternative

In Table 2 we give a simple summary of which test generally dominates for each model and case based on the complete simulation results. In brief, each test dominates for certain models, and in some cases they are comparable. JWW's test generally dominates in models 1, 3, and 4, and for model 7 for larger sample sizes. This applies across error cases, including GARCH errors.

The max-test dominates in models 2, 6, 8 and 9, with strong domination for model 8 (break in variance), and models 2 and 9 (distant nonstationarity). Indeed, JWW's test has only negligible power for models 2, 8 and 9: by construction it cannot detect a break in variance (model 8), and seems incapable of detecting a distant (model 9), or even semi-distant (model 2), form of covariance nonstationarity.

Overall, both tests clearly have merit, and seem to complement each other based on the different cases in which they each excel. Both tests could be applied in practice to glean whether covariance stationarity applies. We do exactly that in the supplemental material for international exchange rates: see Appendix E.

6. Conclusion

We present a max-correlation difference test for testing covariance stationarity in a general setting that allows for nonlinearity and random volatility, and heterogeneity under either hypothesis. Our test exploits a generic orthonormal basis under mild conditions, with Walsh and Haar wavelet function examples. We do not require estimation of an asymptotic covariance matrix, our test can detect a break in variance, and we deliver an asymptotically valid dependent wild bootstrapped p-value. Orthonormal basis based tests direct power toward alternatives implied by basis-specific systematic samples. Thus, by combining bases a power improvement may be achievable. In controlled experiments, however, we find the Walsh basis yields superior properties compared to a composite Haar basis. We leave for future endeavors the question of whether other bases may yet perform better than the Walsh basis for a test of covariance stationarity.

Furthermore, the max-test dominates JWWs in some case, while JWW's dominates in others. The max-test is best capable of delivering sharp empirical size for a nonlinear process and when errors are non-iid; and is particularly suited for detecting distant (large lag) forms of covariance non-stationarity, and a break in variance. The former corroborates findings in Hill and Motegi (2020), who find a max-correlation white noise test strongly dominates Wald and portmanteau tests when there is a distant non-zero correlation. We conjecture this will carry over to other nonstationary models with distant breaks in covariance, but leave this idea for future consideration.

Appendix A: proofs

The following result shows the composite Haar wavelets $\{\psi_k(x)\}$ form a complete orthonormal basis.

Lemma A.1. *a.* $\{\psi_k(x) : 1 \leq k \leq \mathcal{K}_T\}$ *forms a* $\{-1, 1\}$*-valued complete orthonormal basis in* $\mathcal{L}[0, 1)$*; b.* $|\sum_{t=1}^{T} \Psi_k(t)| = O(2^k)$*; c.* $\lim_{T\to\infty} 1/T \sum_{t=1}^{T} \Psi_k(t) = 0$*; d.* $\sum_{t=1}^{T} \Psi_k(t) = 0$ *if* 2^k *is a multiple of* T*.*

Proof.

Claim (a). By construction, for $k = 1, 2, ...,$

$$\psi_k(x) = \sum_{m=0}^{2^{k-1}-1} \psi(2^{k-1}x - m) = \psi(2^{k-1}x) + \psi(2^{k-1}x - 1) + \cdots + \psi(2^{k-1}x - 2^{k-1} + 1),$$

where $\psi(x) \in \{-1, 0, 1\}$, and

$$\psi(2^k x - m) = I\left(\frac{m}{2^k} \leq x < \frac{m + 1/2}{2^k}\right) - I\left(\frac{m + 1/2}{2^k} \leq x < \frac{m + 1}{2^k}\right).$$

For a given couplet (x, k), by mutual exclusivity it follows $\psi(2^k x - m) \in \{-1, 1\}$ for only one $m \in \{0, ..., 2^k - 1\}$. Hence $\psi_k(x) \in \{-1, 1\}$.

Next, by construction of the Haar wavelet functions $\psi(2^k x - m)$:

$$\int_0^1 \psi_k(x)dx = \sum_{m=0}^{2^k-1} \int_0^1 \psi(2^{k-1}x - m)dx = \sum_{m=0}^{2^{k-1}-1} \left(\int_{m/2^{k-1}}^{(m+1/2)/2^{k-1}} dx - \int_{(m+1/2)/2^{k-1}}^{(m+1)/2^{k-1}} dx\right) = 0.$$

Furthermore, $\psi_k(x) \in \{-1, 1\}$ implies $\int_0^1 \psi_k^2(x)dx = 1$. Finally, let $k_1 > k_2$ to yield by the orthogonality of $\{\psi_{k_1,m}(x), \psi_{k_2,m}(x) : k_1 \neq k_2\}$:

$$\begin{aligned}\int_0^1 \psi_{k_1}(x)\psi_{k_2}(x)dx &= \sum_{m_1=0}^{2^{k_1-1}-1} \sum_{m_2=0}^{2^{k_2-1}-1} \int_0^1 \psi(2^{k_1-1}x - m_1)\psi(2^{k_2-1}x - m_2)dx \\ &= \frac{1}{2^{k_1-1/2}2^{k_2-1/2}} \sum_{m_1=0}^{2^{k_1-1}-1} \sum_{m_2=0}^{2^{k_2-1}-1} \int_0^1 \psi(2^{k_1-1}x - m_1)\psi(2^{k_2-1}x - m_2)dx\end{aligned}$$

$$= \sum_{m_1=0}^{2^{k_2-1}-1} \sum_{m_2=0}^{2^{k_2-1}-1} \int_0^1 \psi(2^{k_1-1}x - m_1)\psi(2^{k_2-1}x - m_2)dx$$
$$+ \sum_{m_1=2^{k_2-1}+1}^{2^{k_1-1}-1} \sum_{m_2=0}^{2^{k_2-1}-1} \int_0^1 \psi(2^{k_1-1}x - m_1)\psi(2^{k_2-1}x - m_2)dx$$
$$= \sum_{m=0}^{2^{k_2-1}-1} \int_0^1 \psi(2^{k_1-1}x - m)\psi(2^{k_2-1}x - m)dx$$
$$+ \sum_{m_1=2^{k_2-1}+1}^{2^{k_1-1}-1} \sum_{m_2=0}^{2^{k_2-1}-1} \int_0^1 \psi(2^{k_1-1}x - m_1)\psi(2^{k_2-1}x - m_2)dx = 0.$$

Hence $\{\psi_k(x) : 1 \leq k \leq \mathcal{K}_T\}$ forms a $\{-1, 1\}$-valued orthonormal basis. Completeness follows from completeness of $\{\psi_{k,m}(x) : 1 \leq k \leq \mathcal{K}_T\}$ and the definition $\psi_k(x) \equiv 2^{-(k-1)/2} \sum_{m=0}^{2^{k-1}-1} \psi_{k,m}(x)$.

Claim (b). By construction:

$$\sum_{t=1}^{T} \Psi_k(t) = \sum_{m=0}^{2^k-1} \sum_{t=1}^{T} \psi(2^k (t-1)/T - m) = \sum_{m=0}^{2^k-1} \left(2\left[\left(\frac{m+1/2}{2^k}\right)T\right] - \left[\frac{m}{2^k}T\right] - \left[\left(\frac{m+1}{2^k}\right)T\right]\right) \quad \text{(A.1)}$$

Now use $[aT] - aT \in [-1/2, 1/2]$ $\forall a \in [0, 1]$ to yield for any $m \in \{0, ..., 2^{k-1}\}$:

$$\left|2\left[\left(\frac{m+1/2}{2^k}\right)T\right] - \left[\frac{m}{2^k}T\right] - \left[\left(\frac{m+1}{2^k}\right)T\right]\right|$$
$$\leq \left|2\left(\frac{m+1/2}{2^k}\right)T - \frac{m}{2^k}T - \left(\frac{m+1}{2^k}\right)T\right| + 2\left|\left[\left(\frac{m+1/2}{2^k}\right)T\right] - \left(\frac{m+1/2}{2^k}\right)T\right|$$
$$+ \left|\left[\frac{m}{2^k}T\right] - \frac{m}{2^k}T\right| + \left|\left[\left(\frac{m+1}{2^k}\right)T\right] - \left(\frac{m+1}{2^k}\right)T\right|$$
$$\leq 1 + 1/2 + 1/2 = 2.$$

Therefore $|\sum_{t=1}^{T} \Psi_k(t)| \leq 2^{k+1} = O(2^k)$.

Claim (c). Use (a) with $x_t = (t-1)/T$ and $dx_t = 1/T$ to yield $1/T \sum_{t=1}^{T} \Psi_k(t) = \int_0^1 \psi(2^k x_t - m)dx_t \to \int_0^1 \psi(2^k x - m)dx = 0$ as $T \to \infty$.

Claim (d). The claim follows from identity (A.1), and $[2^{-k}(m+1/2)T] = 2^{-k}(m+1/2)T$ if 2^k divides T (in which case $T/2^k$ is even, hence $2^{-k}(m+1/2)T \in \mathbb{N}$). $\mathcal{QED}$.

The proof of Lemma 3.1 relies on an extension of Assumption 1.b to $\prod_{i=1}^{r} X_{t_i}$ for any r-tuple $\{t_1, ..., t_r\}$, $r \in \mathbb{N}$. This is required here for both couplets $X_t X_{t-h}$ and their cross-products $X_s X_{s-l} X_t X_{t-h}$ for our high dimensional results. See Appendix D.1 of the supplemental material for a proof.

Lemma A.2. *Let* $\max_{1\leq t\leq T} P(|X_t|>c) \leq \varpi \exp\{-c^{\vartheta_1}\mathcal{E}_T^{-\vartheta_2}\}$ *for some* $\varpi > 0$*, any* $\vartheta_1 \geq 2\vartheta_2$ *and* $\vartheta_2 \geq 1$*, and some sequence of constants* $\{\mathcal{E}_T\}$*,* $\liminf_{T\to\infty} \mathcal{E}_T \geq 1$*. It holds that*

$$\max_{1\leq t_1,\ldots,t_r\leq T} P\left(\left|\prod_{i=1}^r X_{t_i}\right| > c\right) \leq r\varpi \exp\left\{-c^{\vartheta_2}\mathcal{E}_T^{-\vartheta_2}\right\}. \tag{A.2}$$

Proof. We prove (A.2) by induction. If $r=1$ then $\max_{1\leq t\leq T} P(|\ X_t| > c) \leq \varpi \exp\{-c^{\vartheta_1}\mathcal{E}_T^{-\vartheta_2}\}$ $\leq \varpi\exp\{-c^{\vartheta_2}\mathcal{E}_T^{-\vartheta_2}\}$ by assumption, given $\vartheta_1 > \vartheta_2 \geq 1$. Now let (A.2) hold for some $r \geq 1$: $\max_{1\leq t_1,\ldots,t_r\leq T} P(|\sqcap_{i=1}^r X_{t_i}| > c) \leq r\varpi \exp\{-c^{\vartheta_2}\mathcal{E}_T^{-\vartheta_2}\}$. The proof is complete if we show (A.2) holds for $r+1$. Young and Bonferroni inequalities yield for any $\vartheta_1 \geq 2\vartheta_2$

$$\begin{aligned}\max_{1\leq t_1,\ldots,t_{r+1}\leq T} P\left(\left|\prod_{i=1}^{r+1} X_{t_i}\right| > c\right) &\leq \max_{1\leq t_1,\ldots,t_{r+1}\leq T} P\left(\frac{1}{2}\left(\prod_{i=1}^r X_{t_i}\right)^2 + \frac{1}{2}X_{t_{r+1}}^2 > c\right)\\ &\leq \max_{1\leq t_1,\ldots,t_r\leq T} P\left(\left|\prod_{i=1}^r X_{t_i}\right| > c^{\frac{1}{2}}\right) + \max_{1\leq t\leq T} P\left(|X_t| > c^{\frac{1}{2}}\right)\\ &\leq r\varpi\exp\left\{-c^{\vartheta_2}\mathcal{E}_T^{-\vartheta_2}\right\} + \varpi\exp\left\{-c^{\vartheta_1/2}\mathcal{E}_T^{-\vartheta_2}\right\}\\ &\leq (r+1)\,\varpi\exp\left\{-c^{\vartheta_2}\mathcal{E}_T^{-\vartheta_2}\right\}. \mathcal{QED}.\end{aligned}$$

In the following we allow for a non-zero mean $E[X_t]\ \forall t$.

Proof of Lemma 3.1. Allowing for a non-zero mean, recall:

$$z_t(h,k) \equiv (X_t - E[X_t])(X_{t+h} - E[X_{t-h}])B_k(t) - \{E[(X_t - E[X_t])(X_{t+h} - E[X_{t-h}])]B_k(t)\}$$

and $\mathcal{Z}_T(h,k) \equiv 1/\sqrt{T}\sum_{t=1}^{T-h} z_t(h,k)$. Here, and in the sequel, let $\{\zeta_t(i), \mathfrak{Z}_T(i)\}_{i=0}^{\mathcal{H}_T\mathcal{K}_T}$ denote $\{z_t(h,k),$ $\mathcal{Z}_T(h,k)\}_{h=0,k=1}^{\mathcal{H}_T,\mathcal{K}_T}$, stacked h-wise over k. For example:

$$\mathfrak{Z}_T(i) = \mathcal{Z}_T(h,k) \text{ with index correspondence } i = (k-1)\mathcal{H}_T + h. \tag{A.3}$$

Thus $\mathfrak{Z}_T(1),\ldots,\mathfrak{Z}_T(\mathcal{H}_T) = \mathcal{Z}_T(1,1),\ldots,\mathcal{Z}_T(\mathcal{H}_T,1)$; $\mathfrak{Z}_T(\mathcal{H}_T+1),\ldots,\mathfrak{Z}_T(2\mathcal{H}_T) = \mathcal{Z}_T(1,2),\ldots,$ $\mathcal{Z}_T(\mathcal{H}_T,2)$; and so on. Define $\sigma_T^2(i) \equiv E\left[\mathfrak{Z}_T^2(i)\right]$ and let $\{\mathbf{Z}_T(i) : T \in \mathbb{N}\}_{i\geq 0}$ be normally distributed $\mathbf{Z}_T(i) \sim N(0,\sigma_T^2(i))$. It suffice to prove the claim for $\mathfrak{Z}_T(i)$.

Under Assumption 1 and Lemma A.2, $\zeta_t(i)$ satisfies Conditions 1-3 in Chang, Chen and Wu (2021). Their Theorem 1 and the mapping theorem therefore imply:

$$\sup_{z\geq 0}\left|P\left(\max_{0\leq i\leq \mathcal{H}_T\mathcal{K}_T} |\mathcal{Z}_T(i)| \leq z\right) - P\left(\max_{0\leq i\leq \mathcal{H}_T\mathcal{K}_T} |\mathbf{Z}_T(i)| \leq z\right)\right| \to 0 \tag{A.4}$$

provided

$$\frac{1}{T^{1/9}}\left[\mathcal{E}_T^{2/3}\{\ln(\mathcal{H}_T\mathcal{K}_T)\}^{(1+2\phi)/(3\phi)} + \mathcal{E}_T(\ln\mathcal{H}_T\mathcal{K}_T)^{7/6}\right] = o(1) \tag{A.5}$$

$$(\ln(\mathcal{H}_T\mathcal{K}_T))^{3-\phi} = o(T^{3\phi}), \tag{A.6}$$

where $\mathcal{E}_T$ is the Assumption 1 exponential scale, $\liminf_{T\to\infty}\mathcal{E}_T \geq 1$, and $\phi > 0$ the mixing coefficient.[10] We need $\mathcal{H}_T = o(T)$ for consistency, and $\mathcal{K}_T = o(T^\kappa)$ for some finite $\kappa > 0$ and $\eta(\mathcal{K}_T) = o(\sqrt{T})$ by Remark 7. Then (A.6) is trivial, and (A.5) becomes $1/T^{1/9}[\mathcal{E}_T^{2/3}\{\ln(T)\}^{(1+2\phi)/(3\phi)} + \mathcal{E}_T(\ln T)^{7/6}] \to 0$. It is easy to show that the first term dominates $\forall\phi > 0$, which reduces to $\mathcal{E}_T = o(T^{1/6}/\{\ln(T)\}^{(1+2\phi)/(2\phi)})$.

Finally, (A.4) implies $\max_{0\leq i\leq\mathcal{H}_T\mathcal{K}_T}|\mathcal{Z}_T(i)| \xrightarrow{d} \max_{i\in\mathbb{N}}|\mathbf{Z}(i)|$ where $\mathbf{Z}(i) \sim N(0, \lim_{T\to\infty}\sigma_T^2(i))$ with $\lim_{T\to\infty}\sigma_T^2(i) < \infty$ shown below. Just note that convergence in distribution follows by construction of $\mathbf{Z}(i)$: $\lim_{T\to\infty}P(\max_{0\leq i\leq\mathcal{H}_T\mathcal{K}_T}|\mathbf{Z}_T(i)| \leq z) = P(\max_{i\in\mathbb{N}}|\mathbf{Z}(i)| \leq z)\ \forall z \geq 0$.

It remains to prove $\lim_{T\to\infty}\sigma_T^2(i) < \infty$. We will prove a uniform bound

$$\max_{1\leq i\leq\mathcal{H}_T\mathcal{K}_T}\sigma_T^2(i) = O(1) \tag{A.7}$$

for future reference. Under Assumption 1 and by measurability, $\{z_t(h,k)\}$ is uniformly (in (h,k)) $\mathcal{L}_r$-bounded for any $r > 2$, and α-mixing with coefficients $\alpha_l = O(\exp\{-l^\phi\})$ for some $\phi > 0$. Then $\{z_t(h,k)\}$ forms a zero-mean $\mathcal{L}_2$-mixingale array with coefficients $\mathring{\alpha}_l = \alpha_l^{1/2-2/r}$ for any $r > 2$ and constants $K||z_t(h,k)||_r$ by Lemma 2.1 in McLeish (1975). The constants satisfy $\max_{\mathcal{H}_T,\mathcal{K}_T}||z_t(h,k)||_r \leq K\max_{\mathcal{H}_T}||X_tX_{t-h}||_r \leq K$ by Minkowsky and Jensen inequalities, $|B_k(t)| = 1$, and Assumption 1.b. Furthermore, $\mathring{\alpha}_l = O(\exp\{-l^{\mathring{\gamma}}\})$ for some $\mathring{\gamma} > 0$. Now use Theorem 1.6 in McLeish (1975) with uniform boundedness of the mixingale constants $\max_{\mathcal{H}_T,\mathcal{K}_T}||z_t(h,k)||_2 \leq K$ to yield (A.7), completing the proof. $\mathcal{QED}$.

The following result proves that we may assume $E[X_t] = 0$ in subsequent proofs to ease notation.

Lemma A.3. *Under Assumption 1, for any $\{\mathcal{E}_T,\mathcal{H}_T,\mathcal{K}_T\}$ with $\mathcal{K}_T = o(T^\kappa)$ for some finite $\kappa > 0$, $\eta(\mathcal{K}_T) = o(\sqrt{T})$ and (3.5):*

$$\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left\{\left(X_t - \bar{X}\right)\left(X_{t-h} - \bar{X}\right) - \left(X_t - \mu\right)\left(X_{t-h} - \mu\right)\right\}B_k(t)\right| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Proof. Write $\tilde{X}_t \equiv X_t - \mu$ and $\hat{X}_t \equiv X_t - \bar{X}$. We have:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left\{\hat{X}_t\hat{X}_{t-h} - \tilde{X}_t\tilde{X}_{t-h}\right\}B_k(t) &= \left(\bar{X}^2 - \mu^2\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}B_k(t) - 2\mu\left(\bar{X} - \mu\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}B_k(t) \\ &\quad - \left(\bar{X} - \mu\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left\{X_{t-h} - \mu + X_t - \mu\right\}B_k(t) \\ &= \mathfrak{A}_T(h,k) + \mathfrak{B}_T(h,k) + \mathfrak{C}_T(h,k).\end{aligned}$$

By Assumption 1 $|1/\sqrt{T}\sum_{t=1}^T B_k(t)| = O(\eta(k)/\sqrt{T})$. Arguments yielding (A.7) identically yield $\bar{X} - \mu = O_p(1/\sqrt{T})$ by Chebyshev's inequality, hence $\bar{X}^2 - \mu^2 = O_p(1/\sqrt{T})$ by the mapping theorem.

[10]Technically (A.5) and (A.6) require $\mathcal{H}_T\mathcal{K}_T + 1$ instead of $\mathcal{H}_T\mathcal{K}_T$ in view of the length of the sequence $\{\mathfrak{Z}_T(i)\}_{i=0}^{\mathcal{H}_T\mathcal{K}_T}$. .Asymptotically, however, the modification is irrelevant.

Therefore, e.g.,

$$\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\left\{\bar{X}^2-\mu^2\right\}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}B_k(t)\right|=O\left(\frac{\max_{\mathcal{K}_T}\eta(k)}{T}\right)=O_p\left(\frac{\eta(\mathcal{K}_T)}{T}\right).$$

Now use $\eta(\mathcal{K}_T)=o(\sqrt{T})$ to yield $\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathfrak{A}_T(h,k)|=o_p(1/\sqrt{T})$. Similarly $\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathfrak{B}_T(h,k)|=o_p(1/\sqrt{T})$.

The remaining term $\mathfrak{C}_T$ is handled by applying arguments in the proof of Lemma 3.1 to deduce for some mean zero Gaussian process $\mathring{\mathbf{Z}}(h,k)\sim N(0,\lim_{T\to\infty}\mathring{\sigma}_T^2(h,k))$ and $\lim_{T\to\infty}\mathring{\sigma}_T^2(h,k)<\infty$:

$$\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\{X_{t-h}-\mu+X_t-\mu\}B_k(t)\right|\xrightarrow{d}\max_{h,k\in\mathbb{N}}|\mathbf{Z}(h,k)|.$$

Hence $\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathfrak{C}_T|=O_p(1/\sqrt{T})$, completing the proof. $\mathcal{QED}$.

Proof of Theorem 3.3. In view of $\sigma_T^2(i)=O(1)$, it follows $\hat{\gamma}_0-\gamma_0=O_p(1/\sqrt{T})$ by Chebyshev's inequality.

Moreover, by construction

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\{X_tX_{t+h}B_k(t)-E\left[X_tX_{t+h}\right]B_k(t)\}=\sqrt{T}(\hat{\gamma}_h^{(k)}-\hat{\gamma}_h)-\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}E\left[X_tX_{t+h}\right]B_k(t). \quad \text{(A.8)}$$

Under covariance stationarity H_0, $|E[X_tX_{t+h}]|<E[X_t^2]<\infty$ for all h and t. The Assumption 1.d basis properties yield for all $\{h,k\}$ and some finite $C>0$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}E[X_tX_{t+h}B_k(t)]=\gamma_h\times\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}B_k(t)=O\left(\frac{\eta(k)}{\sqrt{T}}\right). \quad \text{(A.9)}$$

Hence:

$$\begin{aligned}\sqrt{T}\left(\hat{\rho}_h^{(k)}-\hat{\rho}_h\right)&=\frac{1}{\gamma_0+O_p(1/\sqrt{T})}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\{X_tX_{t+h}B_k(t)-E\left[X_tX_{t+h}\right]B_k(t)\} \quad \text{(A.10)}\\&\quad+\frac{1}{\gamma_0+O_p(1/\sqrt{T})}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}E\left[X_tX_{t+h}\right]B_k(t)\\&=\frac{1}{\gamma_0+O_p(1/\sqrt{T})}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\{X_tX_{t+h}B_k(t)-E\left[X_tX_{t+h}\right]B_k(t)\}+O_p\left(\frac{\eta(k)}{\sqrt{T}}\right),\end{aligned}$$

where the $O_p(\cdot)$ terms do not depend on h. Now exploit $\eta(\mathcal{K}_T)=o(\sqrt{T})$ to yield:

$$\begin{aligned}&\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}\left(\hat{\rho}_h^{(k)}-\hat{\rho}_h\right)\right. \quad \text{(A.11)}\\&\left.\quad-\frac{1}{\gamma_0+O_p(1/\sqrt{T})}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\{X_tX_{t+h}B_k(t)-E\left[X_tX_{t+h}\right]B_k(t)\}\right|=o_p(1).\end{aligned}$$

The claim now follows from Lemma 3.1. $\mathcal{QED}$.

We require a weak convergence result for the bootstrapped correlation difference in order to prove Theorem 4.1. Let $\Rightarrow^p$ denote weak convergence *in probability* on l_∞ (the space of bounded functions) as defined in Giné and Zinn (1990, Section 3). Recall $\{\mathcal{E}_T\}$ is the Assumption 1 exponential moment scale, $\liminf_{T\to\infty}\mathcal{E}_T \geq 1$; the bootstrap index blocks are $\mathfrak{B}_s = \{(s-1)b_T+1,\ldots,sb_T\}$, $s = 1,\ldots,T/b_T$, with block size b_T, $1 \leq b_T < T$, $b_T \to \infty$ and $b_T/T^{1-\iota} \to 0$ for some small $\iota > 0$; ξ_i is iid $N(0,1)$; and $\varphi_t = \xi_s$ if $t \in \mathfrak{B}_s$. Recall

$$\Delta\hat{g}_T^{(dw)}(h,k) \equiv \frac{1}{T}\sum_{t=1}^{T-h}\varphi_t\left\{X_tX_{t+h}B_k(t) - \frac{1}{T}\sum_{t=1}^{T-h}X_tX_{t+h}B_k(t)\right\},$$

and define

$$\mathring{\sigma}_T^2(h,k) \equiv E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\varphi_t\left\{X_tX_{t+h}B_k(t) - \frac{1}{T}\sum_{s=1}^{T-h}E\left[X_sX_{s+h}\right]B_k(s)\right\}\right)^2\right].$$

Recall $\sigma_T^2(h,k) \equiv E[(1/\sqrt{T}\sum_{t=1}^{T-h}z_t(h,k))^2]$ with $z_t(h,k) \equiv \{X_tX_{t+h} - E[X_tX_{t+h}]\}B_k(t)$. Bound (A.7) trivially yields $\lim_{T\to\infty}\sigma_T^2(h,k) < \infty\ \forall(h,k)$. Let $\iota > 0$ be an infinitessimal number that may be different in different places. See Appendix D.2 of the supplemental material for a proof.

Lemma A.4. *Let Assumptions 1 and 2 hold.*

a. Let $\{\mathring{\mathbf{Z}}_T(h,k) : 0 \leq h \leq \mathcal{H}_T, 1 \leq k \leq \mathcal{K}_T\}_{T\leq 1}$ *be a Gaussian process,* $\mathring{\mathbf{Z}}_T(h,k) \sim N(0,\mathring{\sigma}_T^2(h,k))$, *independent of the sample* $\{X_t\}_{t=1}^T$. *For any sequences* $\{\mathcal{E}_T,\mathcal{H}_T,\mathcal{K}_T\}$, *where* $0 \leq \mathcal{H}_T < T-1$, $\mathcal{H}_T = o(T)$, $\mathcal{K}_T = o(T^\kappa)$ *for some finite* $\kappa > 0$, $\eta(\mathcal{K}_T) = o(\sqrt{T})$ *and (A.5) hold:*

$$\sup_{c>0}\left|P\left(\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)\right| \leq c|\{X_t\}_{t=1}^T\right) - P\left(\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\mathring{\mathbf{Z}}_T(h,k)\right| \leq c\right)\right| \xrightarrow{p} 0.$$

b. Let $\{\mathring{\mathbf{Z}}(h,k)\}$ *be an independent copy of the Lemma 3.1 Gaussian process* $\{\mathbf{Z}(k,h) : h,k \in \mathbb{N}\}$, $\mathbf{Z}(h,k) \sim N(0,\lim_{T\to\infty}\sigma_T^2(h,k))$, *independent of the asymptotic draw* $\{X_t\}_{t=1}^\infty$. *For any sequences* $\{b_T,\mathcal{E}_T,\mathcal{H}_T,\mathcal{K}_T\}$, *such that* $0 \leq \mathcal{H}_T < T-1$, $b_T/T^\iota \to \infty$, $b_T = o(T^{1/2-\iota})$, $\mathcal{H}_T = O(T^{1-\iota}/b_T)$, $\mathcal{K}_T = o(T^\kappa)$ *for some finite* $\kappa > 0$, $\eta(\mathcal{K}_T) = o(\sqrt{T})$, *and (A.5) hold:* $\max_{\mathcal{H}_T,\mathcal{K}_T}|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)| \Rightarrow^p \max_{h,k\in\mathbb{N}}|\mathring{\mathbf{Z}}(h,k)|$.

Proof of Theorem 4.1. Operate conditionally on the sample $\mathfrak{X}_T \equiv \{X_t\}_{t=1}^T$. Define max-*covariance* differences $\check{\mathcal{M}}_T \equiv \max_{\mathcal{H}_T,\mathcal{K}_T}|\sqrt{T}(\hat{\gamma}_h^{(k)} - \hat{\gamma}_h)|$ and $\check{\mathcal{M}}_T^{(dw)} \equiv \max_{\mathcal{H}_T,\mathcal{K}_T}|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)|$. Compare this to, e.g., the max-*correlation* difference $\mathcal{M}_T^{(dw)} \equiv \hat{\gamma}_0^{-1}\max_{\mathcal{H}_T,\mathcal{K}_T}|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)|$. Thus, by construction:

$$\hat{p}_{T,M}^{(dw)} \equiv \frac{1}{M}\sum_{i=1}^M I\left(\mathcal{M}_{T,i}^{(dw)} \geq \mathcal{M}_T\right) = \frac{1}{M}\sum_{i=1}^M I\left(\check{\mathcal{M}}_{T,i}^{(dw)} \geq \check{\mathcal{M}}_T\right). \tag{A.12}$$

It suffices to prove the claim for the bootstrapped p-value based on $\check{\mathcal{M}}_T$ and $\check{\mathcal{M}}_{T,i}^{(dw)}$.

By the Glivenko-Cantelli theorem, as $M \to \infty$,

$$\hat{p}_{T,M}^{(dw)} \xrightarrow{p} P\left(\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)\right| \geq \max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}(\hat{\gamma}_h^{(k)}-\hat{\gamma}_h)\right| \mid \mathfrak{X}_T\right). \tag{A.13}$$

Further, $\max_{\mathcal{H}_T,\mathcal{K}_T}|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)| \Rightarrow^p \max_{h,k\in\mathbb{N}}|\mathring{\mathbf{Z}}(h,k)|$ by Lemma A.4, hence

$$\sup_{c>0}\left|P\left(\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)\right| \leq c|\mathfrak{X}_T\right) - P\left(\max_{h,k\in\mathbb{N}}\left|\mathring{\mathbf{Z}}(h,k)\right| \leq c\right)\right| \xrightarrow{p} 0, \tag{A.14}$$

where $\{\mathring{\mathbf{Z}}(h,k) : h,k \in \mathbb{N}\}$ is an independent copy of $\mathbf{Z}(h,k) \sim N(0, \lim_{T\to\infty}\sigma_T^2(h,k))$ from Lemma 3.1, independent of the asymptotic draw $\mathfrak{X}_\infty$. See Giné and Zinn (1990, eq. (3.4)).

Now impose H_0 and define $\bar{F}_T^{(0)}(c) \equiv P(\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathring{\mathbf{Z}}(h,k)| > c)$. Limit (A.14) implies:

$$P\left(\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)\right| \geq \check{\mathcal{M}}_T|\mathfrak{X}_T\right) - P\left(\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\mathring{\mathbf{Z}}(h,k)\right| \geq \check{\mathcal{M}}_T\right) \xrightarrow{p} 0.$$

$[\mathring{\mathbf{Z}}(h,k)]_{h=0,k=1}^{\mathcal{H}_T,\mathcal{K}_T}$ is independent of $\mathfrak{X}_T$, hence:

$$P\left(\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}\Delta\hat{g}_T^{(dw)}(h,k)\right| \geq \check{\mathcal{M}}_T|\mathfrak{X}_T\right) - \bar{F}_T^{(0)}\left(\check{\mathcal{M}}_T\right) \xrightarrow{p} 0. \tag{A.15}$$

$\bar{F}_T^{(0)}$ is continuous by Gaussianicity, thus Lemma 3.1 and Slutsky's theorem yield:

$$\left|\bar{F}_T^{(0)}\left(\check{\mathcal{M}}_T\right) - \bar{F}_T^{(0)}\left(\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathbf{Z}(h,k)|\right)\right| \xrightarrow{p} 0. \tag{A.16}$$

Together, (A.13), (A.15) and (A.16) yield for any sequence of integers $\{M_T\}$, $M_T \to \infty$:

$$\hat{p}_{T,M_T}^{(dw)} = \bar{F}_T^{(0)}\left(\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathbf{Z}(h,k)|\right) + o_p(1). \tag{A.17}$$

Further, $\bar{F}_T^{(0)}(\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathbf{Z}(h,k)|)$ is distributed uniform on $[0,1]$ since $\{\mathring{\mathbf{Z}}(h,k) : h,k \in \mathbb{N}\}$ is an independent copy of $\{\mathbf{Z}(h,k) : h,k \in \mathbb{N}\}$. Thus $P(\hat{p}_{T,M_T}^{(dw)} < \alpha) = P(\bar{F}_T^{(0)}(\max_{\mathcal{H}_T,\mathcal{K}_T}|\mathbf{Z}(h,k)|) < \alpha) + o(1) = \alpha + o(1) \to \alpha$ from (A.17) as required.

Next, impose H_1 defined by (3.14), with drift/basis property (3.9). Thus

$$\frac{1}{T}\sum_{t=1}^{T-h} E[X_t X_{t+h}]B_k(t) \to \int_0^1 c_h(u)\mathcal{B}_k(u)du \neq 0 \text{ for some } h \text{ and } k. \tag{A.18}$$

By the triangle inequality, Lemma 3.1, $\hat{\gamma}_h^{(k)} - \hat{\gamma}_h = 1/T\sum_{t=1}^{T-h} X_t X_{t+h} B_k(t)$ and the definition of $\check{\mathcal{M}}_T$:

$$\begin{aligned}
&\max_{\mathcal{H}_T,\mathcal{K}_T}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} E\left[X_t X_{t+h}\right] B_k(t)\right| \\
&\quad \leq \max_{\mathcal{H}_T,\mathcal{K}_T}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\{X_t X_{t+h} B_k(t) - E\left[X_t X_{t+h}\right] B_k(t)\}\right| + \max_{\mathcal{H}_T,\mathcal{K}_T}\left|\sqrt{T}(\hat{\gamma}_h^{(k)} - \hat{\gamma}_h)\right| = O_p(1) + \check{\mathcal{M}}_T.
\end{aligned}$$

Lemma 3.1 and (A.18) therefore yield:

$$\check{\mathcal{M}}_T \geq \sqrt{T} \max_{\mathcal{H}_T, \mathcal{K}_T} \left| \int_0^1 c_h(u) \mathcal{B}_k(u) du + o(1) \right| + O_p(1) \xrightarrow{p} \infty. \tag{A.19}$$

Finally, combine (A.13), (A.14) and (A.19) to deduce $P(\hat{p}_{T,M_T}^{(dw)} < \alpha) \to 1$ for any $\alpha \in (0,1)$ because:

$$\begin{aligned}
\hat{p}_{T,M_T}^{(dw)} &= P\left(\max_{\mathcal{H}_T, \mathcal{K}_T} \left| \sqrt{T} \Delta \hat{g}_T^{(dw)}(h,k) \right| \geq \max_{\mathcal{H}_T, \mathcal{K}_T} \left| \sqrt{T} (\hat{\gamma}_h^{(k)} - \hat{\gamma}_h) \right| \mid \mathfrak{X}_T \right) + o_p(1) \\
&= P\left(\max_{\mathcal{H}_T, \mathcal{K}_T} |\mathbf{Z}(h,k)| \geq \max_{\mathcal{H}_T, \mathcal{K}_T} \left| \sqrt{T} (\hat{\gamma}_h^{(k)} - \hat{\gamma}_h) \right| \right) + o_p(1) \\
&= \bar{F}_T^{(0)} \left(\max_{\mathcal{H}_T, \mathcal{K}_T} \left| \sqrt{T} (\hat{\gamma}_h^{(k)} - \hat{\gamma}_h) \right| \right) + o_p(1) \xrightarrow{p} 0.
\end{aligned}$$

This proves the claim. $\mathcal{QED}$.

Acknowledgments

The authors thank two expert referees for comments that lead to major improvements of the manuscript.

References

AHMED, N. and RAO, K. R. (1975). *Orthogonal Transformations for Digital Signal Processing*. Springer, New York.

ANDREWS, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59** 817-858.

AUE, A., HÖRMANN, S., HORVÁTH, L. and REIMHERR, M. (2009). Break Detection in the Covariance Structure of Multivariate Time Series Models. *Ann. Stat.* **37** 4046-4087.

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*. Springer.

BUSETT, F. and TAYLOR, A. M. R. (2003). Variance Shifts, Structural Breaks, and Stationarity Tests. *J. Bus. Econ. Stat.* **21** 510-531.

CHANG, J., CHEN, X. and WU, M. (2021). Central Limit Theorems for High Dimensional Dependent Data Technical Report, Dept. of Statistics, University of Illinios, Urbana-Champaign.

CHAREKA, P., CHAREKA, O. and KENNEDY, S. (2006). Locally Sub-Gaussian Random Variables and the Strong Law of Large Numbers. *Atlan. Elec. J. Math.* **1** 75-81.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors. *Ann. Stat.* **41** 2786-2819.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Testing Many Moment Inequalities. Available at arXiv:1312.7614.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors. *Prob. Th. Rel. Fields* **162** 47-70.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central Limit Theorems and Bootstrap in High Dimensions. *Ann. Prob.* **45** 2309-2352.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2019). Inference on Causal and Structural Parameters Using Many Moment Inequalities. *Rev. Econ. Stud.* **86** 1867-1900.

DAHLHAUS, R. (1997). Fitting Time Series Models to Nonstationary Processes. *Ann. Stat.* **25** 1-37.

DAHLHAUS, R. (2009). Local Inference for Locally Stationary Time Series Based on the Empirical Spectral Measure. *J. Econom.* **151** 101-112.

DE JONG, R. M. (1997). Central Limit Theorems for Dependent Heterogeneous Random Variables. *Econom. Th.* **13** 353-367.

DE JONG, R. M. and DAVIDSON (2000). Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices. *Econometrica* **68** 407-423.

DETTE, H., PREUSS, P. and VETTER, M. (2011). A Measure of Stationarity in Locally Stationary Processes with Applications to Testing. *J. Am. Stat. Assoc.* **106** 1113-1124.

DOUKHAN, P. (1994). *Mixing: Properties and Exampls*. Springer, New York.

DWIVEDI, Y. and SUBBA RAO, S. (2011). A Test for Second-Order Stationarity of a Time Series Based on the Discrete Fourier Transform. *J. Time Ser. Anal.* **32** 68-91.

ESCANCIANO, J. C. and LOBATO, I. N. (2009). An Automatic Portmanteau Test for Serial Correlation. *J. Econom.* **151** 140-149.

FRANCQ, C. and ZAKOÏAN, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd ed. Wiley, New York.

GIANETTO, Q. G. and RAISSI, H. (2015). Testing Instantaneous Causality in Presence of Nonconstant Unconditional Covariance. *J. Bus. Econ. Stat.* **33** 46-53.

GINÉ, E. and ZINN, J. (1990). Bootstrapping General Empirical Measures. *Ann. Prob.* **18** 851-869.

GRAY, S. F. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *J. Fin. Econ.* **42** 27-62.

HAAR, A. (1910). Zure Theorie der Orthonormalen Funktionensysteme. *Math. Anal.* **69** 33-371.
HENDRY, D. F. and MASSMANN, M. (2007). Co-Breaking: Recent Advances and a Synopsis of the Literature. *J. Bus. Econ. Stat.* **25** 33-51.
HILL, J. B. and MOTEGI, K. (2020). A Max-Correlation White Noise Test for Weakly Dependent Time Series. *Econom. Th.* **36** 907-960.
INGLOT, T. and LEDWINA, T. (2006). Towards Data Driven Selection of a Penalty Function for Data Driven Neyman Tests. *Lin. Alg. Appl.* **417** 124-133.
JENTSCH, C. and RAO, S. S. (2015). A Test for Second Order Stationarity of a Multivariate Time Series. *J. Econom.* **185** 124-161.
JIN, L., WANG, S. and WANG, H. (2015). A New Non-Parametric Stationarity Test of Time Series in the Time Domain. *J. Roy. Stat. Soc. Ser. B* **77** 893-922.
KREISS, J.-P. and PAPARODITIS, E. (2015). Bootstrapping Locally Stationary Processes. *J. Roy. Stat. Soc. Ser. B* **77** 267-290.
KÜNCH, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Stat.* **17** 1217-1241.
LIU, R. Y. (1988). Bootstrap Procedures under some Non-I.I.D. Models. *Ann. Stat.* **16** 1696-1708.
MCLEISH, D. (1975). A Maximal Inequality and Dependent Strong Laws. *Ann. Prob.* **3** 829-839.
NASON, G. (2013). A Test for Second-Order Stationarity and Approximate Confidence Intervals for Localized Autocovariances for Locally Stationary Time Series. *J. Roy. Stat. Soc. Ser. B* **75** 879-904.
NASON, G., VON SACHS, R. and KROISANDT, G. (200). Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum. *J. Roy. Stat. Soc. Ser. B* **62** 271-292.
NAZAROV, F. (2003). On the Maximal Perimeter of a Convex Set in $\mathbb{R}^n$ with Respect to a Gaussian Measure. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math*, **1807** 69-187. Springer, Berlin.
NEUMANN, M. H. and VON SACHS, R. (1997). Wavelet Thresholding in Anisotropic Function Classes and Application to Adaptive Estimation of Evolutionary Spectra. *Ann. Stat.* **25** 38-76.
PAPARODITIS, E. (2010a). Validating Stationarity Assumptions in Time Series Analysis by Rolling Local Periodograms. *J. Am. Stat. Assoc.* **105** 839-851.
PAPARODITIS, E. (2010b). Validating Stationarity Assumptions in Time Series Analysis by Rolling Local Periodograms. *J. Am. Stat. Assoc.* **105** 839-851.
PERRON, P. (2006). Dealing with Structural Breaks. In *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, (K. Patterson and T. C. Mills, eds.) **1** 278-352. Palgrave Macmillan, Basingstoke, U.K.
PREUSS, P., VETTER, M. . and DETTE, H. (2013). A Test for Stationarity Based on Empirical Processes. *Bernoulli* **19** 2715-2749.
PRIESTLEY, M. B. and SUBBA RAO, T. (1969). A Test for Non-Stationarity of a Time Series. *J. Roy. Stat. Soc. Ser. B* **31** 140-149.
ROSENBLATT, M. (1956). A Central Limit Theorem and A Strong Mixing Condition. *Proc. Nat. Acad. Sci.* **42** 43-47.
SHAO, X. (2010). The Dependent Wild Bootstrap. *J. Am. Stat. Assoc.* **105** 218-235.
SHAO, X. (2011). A Bootstrap-Assisted Spectral Test of White Noise under Unknown Dependence. *J. Econom.* **162** 213-224.
STOFFER, D. S. (1987). Walsh-fourier Analysis of Discrete-Valued Time Series. *J. Time Ser. Anal.* **8** 449-467.
STOFFER, D. S. (1991). Walsh-Fourier Analysis and Its Statistical Applications. *J. Am. Stat. Assoc.* **86** 461-479.
TERÄSVIRTRA, T. (1994). Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *J. Am. Stat. Assoc.* **89** 208-218.
VERSHYNIN, R. (2018). *High-Dimensional Probability*. Cambridge University Press, Cambridge, UK.
VON SACHS, R. and NEUMANN, M. H. (2000). A Wavelet-Based test for Stationarity. *J. Time Ser. Anal.* **21** 597-613.
WALSH, J. L. (1923). A Closed Set of Orthogonal Functions. *Amer. J. Math.* **45** 5-24.
ZHANG, X. and CHENG, G. (2014). Bootstrapping High Dimensional Time Series. Available at arXiv:1406.1037.
ZHANG, D. and WU, W. B. WU (2017). Gaussian Approximation for High Dimensional Time Series. *Ann. Stat.* **45** 1895-1919.