Electronic Journal of Statistics
Vol. 7 (2013) 2241–2297
ISSN: 1935-7524
DOI: 10.1214/13-EJS836

Discriminating between long-range dependence and non-stationarity

Philip Preuß and Mathias Vetter*

Ruhr-Universität Bochum
Fakultät für Mathematik
D-44780 Bochum
Germany
e-mail: philip.preuss@rub.de, mathias.vetter@rub.de

Abstract: This paper is devoted to the discrimination between a stationary long-range dependent model and a non stationary process. We develop a nonparametric test for stationarity in the framework of locally stationary long memory processes which is based on a Kolmogorov-Smirnov type distance between the time varying spectral density and its best approximation through a stationary spectral density. We show that the test statistic converges to the same limit as in the short memory case if the (possibly time varying) long memory parameter is smaller than 1/4 and justify why the limiting distribution is different if the long memory parameter exceeds this boundary. Concerning the latter case the novel FARI(∞) bootstrap is introduced which provides a bootstrap-based test for stationarity which shows good empirical properties if the long memory parameter is smaller than 1/2 which is the usual restriction in the framework of long-range dependent time series. We investigate the finite sample properties of our approach in a comprehensive simulation study and employ the new test in an analysis of two data sets.

AMS 2000 subject classifications: 62M10, 62M15, 62G10.
Keywords and phrases: Bootstrap, empirical spectral measure, goodness-of-fit test, integrated periodogram, locally stationary process, long memory, non stationary process, spectral density.

Received October 2012.

1. Introduction

For many decades one of the leading paradigms in time series analysis is the assumption of stationarity which means that the second-order charateristics of the considered time series are constant over time. One of the prime examples which fits into the framework of stationary processes is the well-known ARMA(p, q) model. Such processes are widely used in applications due to their simplicity and flexibility, and they belong to the class of so called short memory models containing a summable autocovariance function γ.

*This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The authors are very grateful to an Associate Editor and two anonymous referees for their careful reading and for valuable comments that have led to a substantial improvement of the article.

However, many time series in reality exhibit an effect which is known as long-range dependence (or long memory) and which means that γ decays to zero slowly. Usually one has $\gamma(k) \sim Ck^{2d-1}$ as $k \to \infty$ for some $d \in (0, 1/2)$, so in particular the autocovariance function is not absolutely summable. The coefficent d is called long memory parameter, and the most common way to model these kinds of strong dependencies is to employ FARIMA(p, d, q) processes which were introduced in Granger and Joyeux (1980) and Hosking (1981). These long memory extensions of ARMA(p, q) processes are stationary under certain regularity conditions as well. There exists a large literature on long-range dependence in applications, as it occurs e.g. in the modeling of asset volatility, computer network traffic or various other phenomena; see for example Park and Willinger (2000), Henry and Zaffaroni (2002) and Doukhan et al. (2002) for an overview. The assumption of stationarity, however, is always imposed.

More recently, several authors have pointed out that a slow decrease of $\gamma(k)$ might also occur if the true underlying process does actually not possess long memory but is non stationary instead; see Mikosch and Starica (2004), among others. In addition, Starica and Granger (2005) compared the performance of a non stationary model with that of a FARIMA$(1, d, 1)$ and a GARCH$(1, 1)$ process in the framework of volatility forecasting and found out that their non stationary model is leading to superior results. Fryzlewicz et al. (2006) proved that most of the stylized facts which are observed for financial return data can be explained by fitting the simple (but usually non stationary) model

$$X_{t,T} = \sigma(t/T) Z_t, \quad t = 1, \dots, T, \tag{1.1}$$

to the data, where T here and throughout the paper denotes the sample size, $\sigma(\cdot) : [0, 1] \to \mathbb{R}_+$ is a non parametric function and $(Z_t)_t$ is some i.i.d. white noise process. Thus many phenomena in reality can be explained by either fitting a stationary long memory process or a non stationary (short memory) model to the data. A natural question then is how to discriminate between these two approaches.

Although the importance of statistical tests concerning this matter was pointed out by many authors (see e.g. Perron and Qu (2010) or Chen et al. (2010)), there does not exist much research on this topic. Berkes et al. (2006) and Dehling et al. (2013) developed CUSUM and Wilcoxon type tests which discriminate between long-range dependence and changes in mean. While the authors of the first article are testing the null hypothesis that there is no long-range dependence but one change in mean at some unknown point in time (i.e. the alternative corresponds to the case where the process possesses long memory), the latter paper considers the null hypothesis that there is no change in mean but possibly long-range dependence (i.e. the alternative corresponds to the case where there is a change in mean). A similar approach can be found in Sibbertsen and Kruse (2009). However, one can observe many other deviations from stationarity besides changes in mean and it is of particular importance to detect variations in the dependency structure of a given time series as well. There exist some approaches in this area as well, like the one of Lavancier et al. (2011), but, as the articles cited above, they impose rather restrictive conditions either on the type of considered processes or on the class in which the null hypothesis lies. So, in summary, the development of a discrimination procedure in a truly general framework has not been considered yet.

This paper is devoted to the construction of a test for stationarity in the framework of locally stationary long memory processes. The concept of local stationarity became quite famous in recent years, because in contrast to other proposals to model non-stationarity it allows for a meaningful asymptotic theory. Locally stationary processes were introduced by Dahlhaus (1997) and there exist numerous articles which are concerned with estimation techniques or segmentation methods in this framework; see Neumann and von Sachs (1997), Adak (1998), Chang and Morettin (1999), Sakiyama and Taniguchi (2004), Dahlhaus and Polonik (2006), Van Bellegem and von Sachs (2008) or Kreiß and Paparoditis (2011), among others. Articles allowing for long memory effects are rare, however, as only Beran (2009), Palma and Olea (2010) and Roueff and von Sachs (2011) considered parametric and semiparametric estimation.

Similarly, there exist several tests for stationarity in the context of locally stationary models [see for example von Sachs and Neumann (2000), Paparoditis (2009), Paparoditis (2010), Dwivedi and Subba Rao (2010), Dette et al. (2011) and Preuß et al. (2012)], but in all articles long-range dependence is excluded, i.e. these methods cannot be employed for discriminating between long memory and non-stationarity. Our aim is to fill this gap, and for this reason we consider a Kolmogorov-Smirnov type distance which was already discussed in Dahlhaus (2009) and Preuß et al. (2012) to measure deviations from stationarity in the short memory case. Precisely, set

$$E := \sup_{(v,\omega)\in[0,1]^2} |E(v,\omega)|, \tag{1.2}$$

where

$$E(v,\omega) := \frac{1}{2\pi}\Big(\int_0^v\int_0^{\pi\omega} f(u,\lambda)d\lambda du - v\int_0^{\pi\omega}\int_0^1 f(u,\lambda)dud\lambda\Big), \quad (v,\omega)\in[0,1]^2,$$

and $f(u,\lambda)$ denotes the time-varying spectral density. Under the null hypothesis of stationarity $f(u,\lambda)$ does not depend on u and therefore E equals zero. For this reason it is natural to consider an empirical version of the measure in (1.2) in order to construct a test for stationarity.

Even though the literature on empirical spectral processes is quite large in in general (see Dahlhaus (1988), Dahlhaus and Polonik (2009) or Can et al. (2010) among others), those discussing the long memory framework are surprisingly few, even when restricted to the simpler stationary case. To the best of our knowledge, only Kokoszka and Mikosch (1997) have discussed weak convergence of the integrated periodogram to a Gaussian process under stationarity. Our first (and rather probabilistic) goal is therefore to derive the asymptotics of an empirical version $\hat{E}_T(v,\omega)$ of the measure proposed above. Note that neither convergence of the finite dimensional distributions nor asymptotic tightness is self-evident in this context, since the time varying spectral density is typically estimated through a rolling window approach, and it is far from being obvious to what extent the different segments influence each other. Nevertheless, we are able to prove weak convergence of the process $\hat{E}_T(v,\omega)$ to (a discretized version $E_T(v,\omega)$ of) $E(v,\omega)$ at the parametric rate $T^{-1/2}$, but only one if the long memory parameter satisfies $d < 1/4$. This is a natural restriction in this framework (see e.g. Fox and Taqqu (1987) for a similar result on quadratic forms) since the covariances of the finite-dimensional limits contain integrals over the square of the spectral density. These do not exist if the boundary at $1/4$ is exceeded.

This result is obviously of theoretical interest, but it appears unsatisfactory from a statistician's view. Indeed, we obtain a central limit theorem for $\sqrt{T}\sup_{v,\omega}|\hat{E}_T(v,\omega)|$ under the null hypothesis as a consequence, but with a rather complicated dependence structure due to the unknown spectral density and only if $d < 1/4$. Our second main contribution is therefore the invention of the novel FARI(∞) bootstrap for which we are able to prove consistency in the situation above. Interestingly, as it automatically adapts to a switch in the rate of convergence this procedure yields reasonable tests indeed for the entire case of $d < 1/2$ which is the usual assumption in the framework of long-range dependent time series; see for example Berkes et al. (2006) or Giraitis et al. (2012).

The paper is organized as follows. In Section 2 we introduce the necessary notation, whereas we describe the testing procedure in Section 3. The FARI(∞) bootstrap required to obtain asymptotic quantiles of the test statistic is discussed in Section 4, and we investigate the finite sample behaviour of our approach in Section 5. Finally, we defer all proofs to an appendix in Section 6.

2. Locally stationary long memory processes

Locally stationary processes are usually defined via a sequence of stochastic processes $\{X_{t,T}\}_{t=1,\ldots,T}$ which possess a time-varying MA(∞) representation

$$X_{t,T} = \sum_{l=0}^{\infty} \psi_{t,T,l} Z_{t-l}, \quad t = 1,\ldots,T, \tag{2.1}$$

with independent and identically distributed Z_t where $\mathbb{E}(|Z_t|^k) < \infty$ for all $k \in \mathbb{N}$; see Dahlhaus and Polonik (2009). For the coefficents $\psi_{t,T,l}$ we assume that

$$\sup_{t,T} \sum_{l=0}^{\infty} \psi_{t,T,l}^2 < \infty \tag{2.2}$$

is fulfilled which ensures that the process in (2.1) is well defined; see Brockwell and Davis (1991). If the $\psi_{t,T,l}$ are independent of t and T the process $X_{t,T}$ is stationary. However, the coefficents $\psi_{t,T,l}$ depend on t and T in general. To ensure that in this case the process $X_{t,T}$ behaves approximately like a stationary process on a small time interval, it is typically assumed that

$$\sup_{t=1,\dots,T} \sum_{l=0}^{\infty} \big|\psi_{t,T,l} - \psi_l(t/T)\big| = O(1/T) \tag{2.3}$$

holds for twice continuously differentiable functions $\psi_l : [0,1] \to \mathbb{R}$, $l \in \mathbb{Z}$. Different smoothness conditions on the functions $\psi_l(\cdot)$ are imposed in the literature, and in essentially all articles in the framework of local stationarity it is assumed that in addition to (2.2) the condition

$$\sup_{t,T} \sum_{l=0}^{\infty} |\psi_{t,T,l}||l|^{\delta} < \infty \tag{2.4}$$

is satisfied for some $\delta > 0$. This implies $\sup_{t,T} \sum_{h=0}^{\infty} |\mathrm{Cov}(X_{t,T}, X_{t+h,T})| < \infty$, and therefore long memory models are excluded. For this reason we replace (2.4) by a growth condition which is flexible enough to include long-range dependent time series as well. Moreover, instead of using a condition on approximation of the coefficients in an ℓ^1-sense as in (2.3), which is not a natural condition for long memory models whose coefficients are not absolutely summable, we only assume that there exist constants $C > 0$ and $D < 1/2$ such that

$$\sup_{t=1,\dots,T} \Big|\psi_{t,T,l} - \psi_l(t/T)\Big| \le \frac{C}{T}\left(\frac{\log(l)}{l^{1-D}} 1_{\{l\neq 0\}} + 1_{\{l=0\}}\right), \quad \forall l \in \mathbb{N}, \tag{2.5}$$

holds. This condition is obviously more general than (2.3); see e.g. Roueff and von Sachs (2011) for a similar framework.

Assumption 2.1. Suppose we have a sequence of stochastic processes $\{X_{t,T}\}_{t=1,\dots,T}$ which have an MA(∞) representation as in (2.1) with independent and standard normal distributed Z_t such that (2.2) is fulfilled. Furthermore, we assume that (2.5) holds with twice continuously differentiable functions $\psi_l : [0,1] \to \mathbb{R}$ which satisfy the following conditions:

a) There exist twice differentiable functions $a, d : [0,1] \to \mathbb{R}_+$ such that

$$\psi_l(u) = a(u) I(l)^{d(u)-1} + O(I(l)^{D-2}) \tag{2.6}$$

holds uniformly in u as $l \to \infty$, where $D := \sup_{u\in[0,1]} |d(u)| < 1/2$ and $I(x) := |x| \cdot 1_{\{x\neq 0\}} + 1_{\{x=0\}}$.

b) The time varying spectral density

$$f(u,\lambda) := \frac{1}{2\pi}\Big|\sum_{l=0}^{\infty} \psi_l(u) \exp(-i\lambda l)\Big|^2 \tag{2.7}$$

is twice continuously differentiable on $(0,1) \times (0,\pi)$. Furthermore, $f(u,\lambda)$ and all its partial derivatives up to order two are continuous on $[0,1] \times (0,\pi]$.

c) There exists a constant $C \in I\!R$ which is independent of u and λ such that the first and second derivative of the approximating functions $\psi_l(\cdot)$ satisfy

$$\sup_{u\in(0,1)} |\psi_l'(u)| \leq C \log(l) I(l)^{D-1}, \tag{2.8}$$
$$\sup_{u\in(0,1)} |\psi_l''(u)| \leq C \log^2(l) I(l)^{D-1}$$

for $l \neq 0$ and are bounded otherwise. Furthermore, we assume

$$\sup_{u\in(0,1)} |\partial/\partial u \; f(u,\lambda)| \leq C \log(\lambda)/\lambda^{2D},$$
$$\sup_{u\in(0,1)} |\partial^2/\partial u^2 \; f(u,\lambda)| \leq C \log^2(\lambda)/\lambda^{2D}.$$

d) We have

$$\sup_{t,T} |\psi_{t,T,l}| \leq C I(l)^{D-1}. \tag{2.9}$$

To simplify the notation we use $C \in I\!R$ as a universal constant throughout this paper. Note that it is common sense to consider only zero mean processes in this framework since observed data can be easily transformed into data with mean zero. Furthermore, innovations Z_t with a time varying variance $\sigma^2(t/T)$ can be included by choosing other coeffcients $\psi_{t,T,l}$. The assumption of Gaussianity is standard (see Palma and Olea (2010) or Dette et al. (2011)) and only imposed to simplify technical arguments since the proofs are already quite involved in this situation. In addition, the functions $\psi_l(u)$ might have finitely many points of discontinuity without affecting any result stated throughout this article, and we furthermore conjecture that the constraints can be weakened to some kind of condition on the total variation of $\psi_l(u)$ as in Definition 2.1 of Dahlhaus and Polonik (2009).

To obtain an impression for local stationarity, note that the process

$$X_t(u) = \sum_{l=0}^{\infty} \psi_l(u) Z_{t-l} \tag{2.10}$$

is stationary for every $u \in [0,1]$, and that $X_t(t/T)$ serves as an approximation of $X_{t,T}$ in the sense of (2.5). It is easy to see that (2.6) implies

$$|\mathrm{Cov}(X_t(u), X_{t+k}(u))| \sim y_1(u)/k^{1-2d(u)} \quad \text{as } k \to \infty$$

and

$$f(u,\lambda) \sim y_2(u)/\lambda^{2d(u)} \quad \text{as } \lambda \to 0 \tag{2.11}$$

for some functions $y_i(\cdot)$; see the proof of Theorem 3.1 in Palma (2007) for details. This shows that the autocovariance function $\gamma(u,k) = \mathrm{Cov}(X_0(u), X_k(u))$ is not absolutely summable and that the time varying spectral density $f(u,\lambda)$ has a pole in $\lambda = 0$ for every $u \in [0,1]$. If the considered process is stationary then $u \mapsto d(u)$ is independent of u which yields that D equals the long memory parameter d of a stationary time series.

Let us now present an important example which fits into the above framework of locally stationary long memory processes. To this end we define the backshift operator B through $B^k X_t := X_{t-k}$, $k \in \mathbb{N}$, and we set

$$(1-B)^{d(u)} = \sum_{j=0}^{\infty} \binom{d(u)}{j} (-1)^j B^j,$$

just as for the binomial series. The next theorem justifies that both stationary FARIMA(p,d,q) processes and a time-varying extension of them are included in our theoretical framework.

Theorem 2.2. *Consider the system of equations*

$$a(t/T, B) X_{t,T} = b(t/T, B)(1-B)^{-d(t/T)} Z_t, \quad t = 1, \dots, T, \tag{2.12}$$

where the Z_t *are independent and standard normal distributed random variables,* $a_j(\cdot)$ *and* $b_j(\cdot)$ *are twice continuously differentiable functions from* $[0,1]$ *to* $\mathbb{R}$ *with* $a_p(\cdot)$, $b_q(\cdot) \not\equiv 0$, *and* $d(\cdot)$ *is a twice continuously differentiable function from* $[0,1]$ *to* $(0,D)$ *with* $D < 1/2$. *Furthermore, we assume that* $a_0(u) = b_0(u) \equiv 1$ *and that there exists a* $\delta > 0$ *such that for all* $z \in \mathbb{C}$ *with* $|z| \le 1+\delta$ *the condition*

$$\sum_{j=0}^{p} a_j(z) z^j \neq 0 \tag{2.13}$$

is satisfied. Then (2.12) *possesses a locally stationary solution in the sense of Assumption 2.1.*

3. The testing procedure

Let us now come to the development of a test for stationarity in the case of long memory models. We are thus interested in testing the null hypothesis

$$H_0 : f(u,\lambda) \text{ is independent of } u \tag{3.1}$$

against the alternative that there exists an $\lambda \in [0,\pi]$ such that $u \mapsto f(u,\lambda)$ is not independent of u. Our test will be based on empirical versions of the quantities E and $E(v,\omega)$ specified in (1.2), and we see that E vanishes under the null hypothesis while it is positive under the alternative due to the continuity of the spectral density.

In order to obtain an estimator for E we have to define an empirical version of $E(v,\omega)$ at first, and for this reason we require an estimator for $f(u,\lambda)$. We assume without loss of generality that the sample size T can be decomposed as

$T = NM$ where N and M are integers with N even. We then define the local periodogram at the rescaled time point $u \in [0,1]$ by

$$I_N(u,\lambda) := \frac{1}{2\pi N}\Big|\sum_{s=0}^{N-1} X_{\lfloor uT\rfloor - N/2+1+s,T}\exp(-i\lambda s)\Big|^2$$

[see Dahlhaus (1997)], where we have set $X_{j,T} = 0$, if $j \notin \{1,\dots,T\}$. This is the usual periodogram computed from the observations $X_{\lfloor uT\rfloor - N/2+1,T}, \dots$, $X_{\lfloor uT\rfloor + N/2,T}$. It can be shown that the quantity $I_N(u,\lambda)$ is an asymptotically unbiased estimator for the local spectral density if $N \to \infty$ and $N = o(T)$. However, $I_N(u,\lambda)$ is not consistent just as the usual periodogram.

An empirical version of $E(v,\omega)$ is now constructed by replacing the integral by its Riemann sum and substituting the time varying spectral density $f(u,\lambda)$ by its (asymptotically) unbiased estimator $I_N(u,\lambda)$. In other words, we define an estimator for $E(v,\omega)$ by

$$\hat{E}_T(v,\omega) := \frac{1}{T}\sum_{j=1}^{\lfloor vM\rfloor}\sum_{k=1}^{\lfloor \omega\frac{N}{2}\rfloor} I_N(u_j,\lambda_k) - \frac{\lfloor vM\rfloor}{M}\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor \omega\frac{N}{2}\rfloor} I_N(u_j,\lambda_k), \qquad (3.2)$$

where $u_j := t_j/T := (N(j-1)+N/2)/T$ and $\lambda_k := 2\pi k/N$ with $j = 1,\dots,M$ and $k = 1,\dots,N/2$. Note that in this procedure the T observations are divided into M intervals with length N and that the u_j correspond to the midpoints of these intervals in rescaled time. The λ_k are the usual Fourier frequencies. We then set

$$E_T(v,\omega) := \frac{1}{T}\sum_{j=1}^{\lfloor vM\rfloor}\sum_{k=1}^{\lfloor \omega\frac{N}{2}\rfloor} f(u_j,\lambda_k) - \frac{\lfloor vM\rfloor}{M}\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor \omega\frac{N}{2}\rfloor} f(u_j,\lambda_k),$$

which is the Riemann sum approximation of $E(v,\omega)$, and consider the empirical spectral process

$$\hat{G}_T(v,\omega) := \hat{E}_T(v,\omega) - E_T(v,\omega), \quad v,\omega \in [0,1].$$

Alternatively, an estimator for the time-varying spectral density could be based on the pre-periodogram

$$J_T(u,\lambda) := \frac{1}{2\pi}\sum_{k:1\le\lfloor uT+1/2\pm k/2\rfloor\le T} X_{\lfloor uT+1/2+k/2\rfloor}X_{\lfloor uT+1/2-k/2\rfloor}\exp(-i\lambda k),$$

which was introduced by Neumann and von Sachs (1997) and further discussed in Dahlhaus (2009) and Preuß et al. (2012) in the short memory context. The main advantage of the pre-periodogram is that no specification of a tuning parameter such as N is necessary. However, as discussed in an extensive simulation study in Preuß et al. (2012), a test based on this concept leads to a substantial loss in power, which is why we restrict ourselves to local periodograms in the following.

The following theorem specifies the asymptotic properties of the process $(\hat{G}_T(v,\omega))_{v,\omega}$ in the case $D < 1/4$, in which the empirical spectral process converges at the parametric rate $T^{-1/2}$ to a mean zero Gaussian process. Note that the results hold both under the null hypothesis and the alternative, and throughout this paper the symbol $\Rightarrow$ denotes weak convergence in $L^\infty([0,1]^2)$.

Theorem 3.1. *Suppose that Assumption 2.1 with $D < 1/4$ is satisfied and let*

$$N \to \infty, \quad N/T \to 0. \tag{3.3}$$

Then as $T \to \infty$ we have

$$\sqrt{T}(\hat{G}_T(v,\omega) - C_T(v,\omega))_{(v,\omega)\in[0,1]^2} \Rightarrow (G(v,\omega))_{(v,\omega)\in[0,1]^2},$$

where $(G(v,\omega))_{(v,\omega)\in[0,1]^2}$ is a Gaussian process with mean zero and covariance structure

$$\operatorname{Cov}(G(v_1,\omega_1), G(v_2,\omega_2)) = \frac{1}{2\pi}\int_0^1 \int_0^{\pi \min(\omega_1,\omega_2)} (1_{[0,v_1]}(u) - v_1)(1_{[0,v_2]}(u) - v_2) \times f^2(u,\lambda) d\lambda du.$$

$C_T(v,\omega)$ denotes a bias term which equals zero if the functions $\psi_l(u)$ are independent of u for all $l \in \mathbb{Z}$ and which is some $O(N^2/T^2 + \log(N)/N^{1-2D})$, uniformly in v,ω, otherwise.

Even under the alternative the bias term above is negligible for $D < 1/6$, at least for a suitable choice of N. This is why it does not appear in the related result in Preuß et al. (2012). More interesting for us is the behaviour under (3.1), however. In this case we have $C_T(v,\omega) = E_T(v,\omega) = 0$ for all v,ω,T. Thus Theorem 3.1 implies

$$(\sqrt{T}\hat{E}_T(v,\omega))_{(v,\omega)\in[0,1]^2} \Rightarrow (G(v,\omega))_{(v,\omega)\in[0,1]^2}$$

under the null hypothesis which yields

$$\sqrt{T} \sup_{(v,\omega)\in[0,1]^2} |\hat{E}_T(v,\omega)| \xrightarrow{D} \sup_{(v,\omega)\in[0,1]^2} |G(v,\omega)|.$$

An asymptotic level α test is then given by rejecting (3.1) whenever $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2} |\hat{E}_T(v,\omega)|$ exceeds the (in principle unknown) $(1-\alpha)$ quantile of the distribution of the random variable $\sup_{(v,\omega)\in[0,1]^2} |G(v,\omega)|$. To obtain consistency of the test, note that $E_T(v,\omega) \geq C$ for some $v,\omega \in [0,1]$ and T large enough, if we are under the alternative. Since Theorem 3.1 implies $|\hat{E}_T(v,\omega) - E_T(v,\omega)| \to 0$ in probability for this specific (v,ω), it follows that $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2} |\hat{E}_T(v,\omega)|$ blows up to infinity (in probability).

Even under the null hypothesis the distribution of the limiting distribution depends in a complicated way on the unknown spectral density. For this reason, we introduce the FARI(∞) bootstrap in the next section and prove that it can be employed to approximate the distribution of $\sqrt{T}\sup_{(v,\omega)\in[0,1]^2} |\hat{E}_T(v,\omega)|$ which implies a test for stationarity in the case of Theorem 3.1.

The restriction $D < 1/4$ in Theorem 3.1 is necessary since $f^2(u, \lambda)$ in the asymptotic variance is not integrable anymore if $D \geq 1/4$ due to (2.11). In fact, in the latter case the rate of convergence is different to $T^{-1/2}$ and the calculation of the corresponding variance (not to even mention higher moments) becomes extremely messy. However, we are able to prove tightness of the process $(\beta_T^{1/2}(\hat{G}_T(v,\omega) - C_T(v,\omega)))_{(v,\omega)\in[0,1]^2}$ in general (see the auxiliary Theorem 6.1 in the Appendix and its use for establishing tightness of several other variables), where

$$\beta_T = \begin{cases} T, & D < 1/4, \\ T/\log N, & D = 1/4, \\ TN^{1-4D}, & D > 1/4. \end{cases} \tag{3.4}$$

We conjecture that a central limit theorem holds with the rate specified above, but we dispense with the precise statement of such claims. We see from the simulation results in Section 5 that the FARI(∞) bootstrap possesses good empirical properties in this situation as well, even though a formal proof that the test keeps the exact asymptotic level relies on a central limit theorem which we do not provide. Note that we require D to be smaller than $1/2$ in any case, as this is the usual restriction in this framework since for example FARIMA(p, d, q) models are not stationary anymore if $D \geq 1/2$.

4. Bootstrapping the test statistic

In this section we introduce a bootstrap procedure which approximates the distribution of $\beta_T^{1/2} \sup_{(v,\omega)\in[0,1]^2} |\hat{E}_T(v,\omega)|$ in the case $D < 1/2$. We call our procedure the FARI(∞) bootstrap as it extends the AR(∞) bootstrap of Kreiß (1988) to the long memory situation. While the AR(∞) bootstrap works by choosing a $p = p(T) \in \mathbb{N}$ and then fitting an AR(p) model to the data, the FARI(∞) bootstrap fits an FARIMA$(p, d, 0)$ model to the data where in both cases $p = p(T)$ grows to infinity as T gets larger. We will describe this method in more detail later and state now the main technical assumptions which will be required.

Assumption 4.1. For the stationary process X_t with strictly positive spectral density $\lambda \mapsto \int_0^1 f(u, \lambda)du$, there exists a $0 < D < 1/2$ such that the process $Y_t = (1 - B)^D X_t$ possesses an AR(∞)-representation, i.e.

$$Y_t = \sum_{j=1}^{\infty} a_j Y_{t-j} + \sigma Z_t, \tag{4.1}$$

where the $(Z_t)_{t\in\mathbb{Z}}$ denote independent standard normal distributed random variables, $\sigma^2 > 0$, $1 - \sum_{j=1}^{\infty} a_j z^j \neq 0$ for $|z| \leq 1$ and

$$\sum_{j=1}^{\infty} |a_j| j^7 < \infty. \tag{4.2}$$

The proof of Theorem 2.2 reveals that X_t has the representation

$$X_t = \sum_{l=0}^{\infty} \tilde{\psi}_l Z_{t-l} \tag{4.3}$$

with coefficents $\tilde{\psi}_l$ satisfying Assumption 2.1, and the aim of the bootstrap procedure is to reproduce the behaviour of the previous test statistic in case the process X_t was observed. Note that under the null hypothesis X_t basically equals $X_{t,T}$, that D is the corresponding long memory parameter and that $\tilde{\psi}_l$ is close to $\psi_{l,t,T}$ in the sense of (2.5).

We start by choosing some $p = p(T) \in \mathbb{N}$, estimating D through some $\hat{D}$ and then fitting an AR(p) model to the process Y_t from (4.1), i.e. estimating

$$(a_{1,p}, \ldots, a_{p,p}) = \underset{b_{1,p},\ldots,b_{p,p}}{\operatorname{argmin}} \mathbb{E}\Big(Y_t - \sum_{j=1}^{p} b_{j,p} Y_{t-j}\Big)^2.$$

We then consider the approximating process $Y_t^{AR}(p)$ which is defined through

$$Y_t^{AR}(p) = \sum_{j=1}^{p} a_{j,p} Y_{t-j}^{AR}(p) + Z_t^{AR}(p), \tag{4.4}$$

where $Z_t^{AR}(p)$ is a Gaussian white noise process with mean zero and variance $\sigma_p^2 = \mathbb{E}(Y_t - \sum_{j=1}^{p} a_{j,p} Y_{t-j})^2$. The idea is that for $p = p(T) \to \infty$ the process $Y_t^{AR}(p)$ is close to the process Y_t and therefore $(1-B)^{-D} Y_t^{AR}(p)$ is close to the stationary process X_t whose spectral density is given through $\lambda \mapsto \int_0^1 f(u,\lambda) du$ as well.

So if we observe the data $X_{1,T}, \ldots, X_{T,T}$, the FARI(∞) bootstrap precisely works as follows:

1) Choose $p = p(T) \in \mathbb{N}$ and calculate $\hat{\theta}_{T,p} = (\hat{D}, \hat{\sigma}_p^2, \hat{a}_{1,p}, \ldots, \hat{a}_{p,p})$ as the minimizer of

$$\frac{1}{T} \sum_{k=1}^{T/2} \Big(\log f_{\theta_p}(\lambda_{k,T}) + \frac{I_T(\lambda_{k,T})}{f_{\theta_p}(\lambda_{k,T})} \Big)$$

where $\lambda_{k,T} = 2\pi k/T$ for $k = 1, \ldots, T/2$, $I_T(\lambda) = \frac{1}{2\pi T} |\sum_{t=1}^{T} X_{t,T} \exp(-i\lambda t)|^2$ is the usual periodogram for stationary processes and

$$f_{\theta_p}(\lambda) = \frac{|1 - \exp(-i\lambda)|^{-2D}}{2\pi} \times \frac{\sigma_p^2}{|1 - \sum_{j=1}^{p} a_{j,p} \exp(-i\lambda j)|^2}$$

is the spectral density of a stationary FARIMA$(p, D, 0)$ model which we want to fit. Note that the estimator $\hat{\theta}_{T,p}$ is the classical Whittle estimator of a stationary process; see Whittle (1951).

2) Simulate a pseudo-series $(Y_t^*)_{t\in\mathbb{Z}}$ according to the model

$$Y_t^* = \sum_{j=1}^{p} \hat{a}_{j,p} Y_{t-j}^* + \hat{\sigma}_p Z_t^*,$$

where the Z_t^* are independent standard normal distributed random variables. Note that in practice it is not possible to simulate such an infinite series. We comment further on this issue at the beginning of Section 5.

3) Create the pseudo-series $X_{1,T}^*, \dots, X_{T,T}^*$ by calculating $X_{i,T}^* = (1-B)^{-\hat{D}} Y_{i,T}^*$ and compute $\hat{E}_T^*(v,\omega)$ in the same way as $\hat{E}_T(v,\omega)$ but with the original observations $X_{1,T}, \dots, X_{T,T}$ replaced by the bootstrap replicates $X_{1,T}^*, \dots, X_{T,T}^*$.

Our goal now is to prove consistency of the FARI(∞) bootstrap which is concerned with the series $X_{t,T}^*$. Some technical assumptions on rates regarding p and $\hat{\theta}_{T,p}$ are necessary which are standard in the framework of an $AR(\infty)$ bootstrap; see for example Berg et al. (2010) or Kreiß et al. (2011).

Assumption 4.2. i) We have $p = p(T) \in [p_{min}(T), p_{max}(T)]$ with $p_{min}(T) \to \infty$, where also

$$p_{max}^{11}(T)\log(T)^2/T \le C \quad \text{and} \quad \sqrt{T} p_{min}(T)^{-10} \to 0.$$

ii) The condition

$$\max_{1\le p\le p_{\max}(T)} ||\hat{\theta}_{T,p} - \theta_p|| = O_P(\sqrt{p_{\max}(T)/T}), \tag{4.5}$$

holds, where $\theta_p = (D, \sigma_p^2, a_{1,p}, \dots, a_{p,p})$ denotes the vector of the true parameters.

Note that a rigorous proof of condition (4.5) is missing in the case of (locally) stationary long memory models. At least, we know from Theorem 2.1 in Hannan and Kavalieris (1986) that the even stronger relation

$$\max_{1\le j\le p_{\max}(T)} ||\hat{\theta}_{T,p} - \theta_p||_\infty = O_P(\sqrt{\log(T)/T})$$

holds true for linear short memory models [see also the discussion on Assumption 3 of Berg et al. (2010)], and Fox and Taqqu (1986) show that the parameters of a stationary long memory model with a finite number of parameters can be estimated with rate $T^{-1/2}$. These examples indicate that (4.5) holds in our specific class of long memory processes as well. However, determining a general class of processes for which such a conditions holds is an open problem. Similarly under the alternative: In this case, Dahlhaus (1997) proves that in a (short memory) locally stationary model the Yule-Walker estimator converges at rate $T^{-1/2}$ to the parameters of the best stationary approximation in the sense above. Again, we need an extension to the long memory context in the sense of Fox and Taqqu (1986) and a result on the behaviour for growing p as in Hannan and Kavalieris (1986) which are both not available so far. A detailed treatment of this conjecture is beyond the scope of the paper, however.

Let us mention some implications: Assumption 4.1 together with Lemma 2.3 of Kreiß et al. (2011) yields that there exists a $p_0 \in \mathbb{N}$ such that for all $p \geq p_0$ the approximating process $Y_t^{AR}(p)$ defined in (4.4) possesses an MA(∞) representation

$$Y_t^{AR}(p) = \sum_{l=0}^{\infty} c_{l,p} Z_{t-l}^{AR}(p).$$

In order to obtain such an MA(∞) representation, the authors use the fact that the characteristic polynomial of the autoregressive part has no zeroes inside the unit disc. Therefore, employing condition (4.5) we obtain a similar form of the fitted AR(p) process Y_t^*, namely

$$Y_t^* = \sum_{l=0}^{\infty} \hat{c}_{l,p} Z_{t-l}^*, \tag{4.6}$$

with a probability tending to one as T increases. Note also that the additional condition

$$\sum_{l=0}^{\infty} |c_{l,p}| l^7 \leq C < \infty \tag{4.7}$$

holds, due to (4.2) and Lemma 2.4 of Kreiß et al. (2011). We can use these relations to investigate the properties of an $MA(\infty)$ representation of the bootstrap replicates $X_{t,T}^*$. If $\hat{D} > 0$, a Taylor expansion yields

$$(1-z)^{-\hat{D}} = \sum_{l=0}^{\infty} \hat{\eta}_l z^l \quad \text{ with } \hat{\eta}_l := \frac{\Gamma(l+\hat{D})}{\Gamma(\hat{D})\Gamma(l+1)}$$

for $l \in \mathbb{N}$; see (6.37) with $d(u)$ replaced by $\hat{D}$. Otherwise, for $\hat{D} = 0$ we have $\hat{\eta}_l = 1_{\{l=0\}}$. Using this expansion and (4.6) we obtain

$$X_{t,T}^* = (1-B)^{-\hat{D}} Y_t^* = \sum_{l=0}^{\infty} \hat{\psi}_{l,p} Z_{t-l}^*, \tag{4.8}$$

where the parameters $\hat{\psi}_{l,p}$ are given through the relation

$$\hat{\psi}_{l,p} = \sum_{k=0}^{l} \hat{c}_{k,p} \hat{\eta}_{l-k}; \tag{4.9}$$

see for example the proof of Lemma 3.2 in Kokoszka and Taqqu (1995).

Recall that the $X_{t,T}^*$ are designed as replicates of the stationary process X_t with $MA(\infty)$ representation (4.3). Once we show consistency of the FARI(∞)

bootstrap later, we will naturally use similar arguments as in the proof of Theorem 3.1. For this reason we require the coefficents $\hat{\psi}_{l,p} - \tilde{\psi}_l$ to satisfy conditions which are similar to the conditions on the true coefficents as stated in Assumption 2.1. Note that the coefficents $\hat{\psi}_{l,p} - \tilde{\psi}_l$ do not depend on the rescaled time u. Therefore all conditions but (2.6) in Assumption 2.1 are automatically fulfilled and the following lemma ensures that we obtain a condition similar to (2.6) as well.

Lemma 4.3. *Suppose that the Assumptions 2.1, 4.1 and 4.2 are satisfied. Then we have*

$$|\hat{\psi}_{l,p} - \tilde{\psi}_l| l^{1-\max(\hat{D},D)} = O_P(p^5/\sqrt{T}), \quad \textit{uniformly in } p, l, \hat{D}, D.$$

Let us now state the formal bootstrap test for stationarity. Empirical quantiles of $\sup_{(v,\omega)\in[0,1]^2} |\hat{E}_T(v,\omega)|$ are obtained by calculating

$$\hat{F}^*_{T,i} := \sup_{(v,\omega)\in[0,1]^2} |\hat{E}^*_{T,i}(v,\omega)| \quad \text{for } i = 1, \dots, B,$$

where $\hat{E}^*_{T,1}(v,\omega), \dots, \hat{E}^*_{T,B}(v,\omega)$ are the B bootstrap replicates of $\hat{E}_T(v,\omega)$. We then reject the null hypothesis, whenever

$$\sup_{(v,\omega)\in[0,1]^2} |\hat{E}_T(v,\omega)| > (\hat{F}^*_T)_{T,\lfloor(1-\alpha)B\rfloor}, \tag{4.10}$$

where $(\hat{F}^*_T)_{T,1}, \dots, (\hat{F}^*_T)_{T,B}$ denotes the order statistic of $\hat{F}^*_{T,1}, \dots, \hat{F}^*_{T,B}$. Note that there is no need to standardise either side with the (in principle unknown) factor $\beta_T^{1/2}$ from (3.4).

In order to explain why this bootstrap procedure works, we have to introduce approximations of $\hat{E}_T(v,\omega)$ and $\hat{E}^*_T(v,\omega)$. First, if we replace $X_{t,T}$ in the definition of $\hat{E}_T(v,\omega)$ by $X_t(t/T)$ from (2.10), we denote the resulting process with $\hat{E}_{T,a}(v,\omega)$. Similarly, we set

$$X^*_{t,T,a} = \sum_{l=0}^{\infty} \tilde{\psi}_l Z^*_{t-l}, \tag{4.11}$$

where the Z^*_t are the innovations from part 2) above. We then define $\hat{E}^*_{T,a}(v,\omega)$ in the same way as $\hat{E}^*_T(v,\omega)$, but with the bootstrap series $X^*_{t,T}$ replaced by $X^*_{t,T,a}$.

Lemma 4.4. *Let the Assumptions 2.1, 4.1 and 4.2 be fulfilled and choose N in such a way that $N \sim cT^\kappa$ for some $0 < \kappa < 1$ and some $c > 0$. If the null hypothesis (3.1) holds, we have*

a)

$$\sup_{(v,\omega)\in[0,1]^2} |\hat{E}_{T,a}(v,\omega)| \overset{\mathcal{D}}{=} \sup_{(v,\omega)\in[0,1]^2} |\hat{E}^*_{T,a}(v,\omega)|,$$

b)

$$\beta_T^{1/2}\Big(\sup_{(v,\omega)\in[0,1]^2}|\hat{E}_T(v,\omega)| - \sup_{(v,\omega)\in[0,1]^2}|\hat{E}_{T,a}(v,\omega)|\Big) = o_P(1).$$

Also, both under the null hypothesis and the alternative we have

c)

$$\beta_T^{1/2}\Big(\sup_{(v,\omega)\in[0,1]^2}|\hat{E}_T^*(v,\omega)| - \sup_{(v,\omega)\in[0,1]^2}|\hat{E}_{T,a}^*(v,\omega)|\Big) = o_P(1),$$

d)

$$(\beta_T^{1/2}\hat{E}_T^*(v,\omega))_{(v,\omega)\in[0,1]^2} \text{ is tight.}$$

It has been indicated in Paparoditis (2010) that Lemma 4.4 a)–c) are sufficient to prove that the test constructed in (4.10) has exact asymptotic level α, but an important ingredient in its proof is the weak convergence of $\sup_{(v,\omega)\in[0,1]^2}\beta_T^{1/2}|\hat{E}_T(v,\omega)|$ with a continuous limit distribution. This result is only available for $D < 1/4$.

Although we cannot use Lemma 4.4 to show that the bootstrap test keeps the exact asymptotic level α even for a general $D < 1/2$, a conservative test based on it can be constructed as well.

Theorem 4.5. *Suppose that the assumptions of Lemma 4.4 are satisfied and let $Q_T^*(1-\alpha)$ denote the $1-\alpha$ quantile of the bootstrap statistic $\sup_{(v,\omega)\in[0,1]^2}\beta_T^{1/2}|\hat{E}_T^*(v,\omega)|$.*

a) If $D < 1/4$, under the null hypothesis we have

$$P\Big(\sup_{(v,\omega)\in[0,1]^2}\beta_T^{1/2}|\hat{E}_T(v,\omega)| \le Q_T^*(1-\alpha)\Big) \to 1-\alpha.$$

b) Let $\delta > 0$ be arbitrary. Then, under the null hypothesis we have

$$\liminf_{T\to\infty} P\Big(\sup_{(v,\omega)\in[0,1]^2}\beta_T^{1/2}|\hat{E}_T(v,\omega)| \le Q_T^*(1-\alpha)+\delta\Big) \ge 1-\alpha.$$

Consistency of the test in (4.10) is granted from Lemma 4.4 d) in any case, since each bootstrap statistic $\sup_{(v,\omega)\in[0,1]^2}|\hat{E}_T^*(v,\omega)|$ converges to zero then, while $\sup_{(v,\omega)\in[0,1]^2}|\hat{E}_T(v,\omega)|$ becomes larger than some positive constant under the alternative due to Theorem 6.1 a), b) and (2.11).

5. Finite sample properties

Our aim now is to demonstrate how the test for stationarity performs in finite sample situations. Since the proposed decision rule (4.10) depends on the choice of N in the estimation of the Kolmogorov-Smirnov type distance and furthermore on the selection of the AR parameter p in the bootstrap procedure, we start by discussing how we choose both parameters. We then investigate the size and power of our test where all reported results are based on 200 bootstrap replications and 1000 simulation runs. Finally we apply our test to two data sets, one regarding tree ring data and one containing S&P 500 returns.

5.1. Choice of the parameters N and p

Although the proposed method does not show much sensitivity with respect to different choices of the AR parameter, we select p throughout this section as the minimizer of the AIC criterion dating back to Akaike (1973), which is defined by

$$\hat{p} = \text{argmin}_p \frac{2\pi}{T} \sum_{k=1}^{T/2} \Big(\log f_{\hat{\theta}(p)}(\lambda_{k,T}) + \frac{I_T(\lambda_{k,T})}{f_{\hat{\theta}(p)}(\lambda_{k,T})} \Big) + p/T$$

in the context of stationary processes due to Whittle (1951). Here, $f_{\hat{\theta}(p)}$ is the spectral density of the fitted stationary FARIMA$(p, D, 0)$ process and I_T is the usual stationary periodogram; see step 1) in the description of the FARI(∞) bootstrap. Therefore we focus in the following discussion on a sensitivity analysis of the test (4.10) with respect to different choices of N. We will see that the particular choice of that tuning parameter has typically very little influence on the outcome of the test under the null hypothesis while it can change the power substantially under certain alternatives.

Note further that in practice it is not feasible to create an infinite series $(Y_t^*)_{t\in\mathbb{Z}}$ as described in step 2) of the FARI(∞) procedure. In order to circumvent this problem during the simulation study we follow a pragmatic approach and replace step 2) by

2*) Calculate $Y_{t,T} = (1-B)^{\hat{D}} X_t^{(b)}$, where $X_t^{(b)} = X_{t,T}$ for $t = 1, \ldots, T$ and $X_t^{(b)} = 0$ for $t \leq 0$. Then simulate a pseudo-series $Y_{1,T}^*, \ldots, Y_{T,T}^*$ according to

$$Y_{t,T}^* = Y_{t,T};\ t = 1, \ldots, p, \quad Y_{t,T}^* = \sum_{j=1}^{p} \hat{a}_{j,p} Y_{t-j,T}^* + \hat{\sigma}_p Z_t^*;\ p < t \leq T,$$

where the Z_t^* are independent standard normal distributed random variables.

5.2. Size of the test

In order to study the approximation of the nominal level, we consider the FARIMA$(1, d, 1)$ model

$$(1 - \phi B)(1 - B)^d X_t = (1 + \theta B) Z_t \tag{5.1}$$

for independent and standard Gaussian Z_t and present the results for different values of ϕ, θ and d. To be more precise, we simulate

$$(1 - \phi B)(1 - B)^d X_t = Z_t \tag{5.2}$$

and

$$(1 - B)^d X_t = (1 + \theta B) Z_t \tag{5.3}$$

for $d \in \{0.2, 0.4\}$ and $\phi, \theta \in \{-0.9, -0.5, 0, 0.5, 0.9\}$. The corresponding results for $d = 0.2$ are depicted in Tables 1 and 2 for the models (5.2) and (5.3), respectively. In the latter case we observe a precise approximation of the nominal

TABLE 1
Rejection probabilities of the test (4.10) under the null hypothesis. The data was generated according to model (5.1) with $d = 0.2$, $\theta = 0$ and different values for ϕ

			$\phi = -0.9$		$\phi = -0.5$		$\phi = 0$		$\phi = 0.5$		$\phi = 0.9$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
128	16	8	0.131	0.179	0.064	0.098	0.054	0.087	0.078	0.122	0.104	0.17
128	8	16	0.129	0.167	0.069	0.11	0.056	0.102	0.086	0.127	0.095	0.151
256	32	8	0.093	0.129	0.056	0.099	0.039	0.072	0.051	0.083	0.087	0.152
256	16	16	0.069	0.107	0.057	0.088	0.041	0.086	0.068	0.124	0.08	0.118
256	8	32	0.067	0.112	0.046	0.093	0.046	0.09	0.077	0.118	0.051	0.096
512	64	8	0.051	0.099	0.047	0.086	0.039	0.087	0.031	0.07	0.062	0.108
512	32	16	0.058	0.109	0.048	0.097	0.043	0.087	0.051	0.1	0.077	0.14
512	16	32	0.056	0.109	0.046	0.085	0.062	0.115	0.066	0.112	0.054	0.122
512	8	64	0.052	0.092	0.05	0.1	0.033	0.086	0.065	0.118	0.041	0.091

TABLE 2
Rejection probabilities of the test (4.10) under the null hypothesis. The data was generated according to model (5.1) with $d = 0.2$, $\phi = 0$ and different values for θ

			$\theta = -0.9$		$\theta = -0.5$		$\theta = 0.5$		$\theta = 0.9$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
128	16	8	0.075	0.124	0.066	0.124	0.061	0.106	0.059	0.092
128	8	16	0.064	0.112	0.058	0.101	0.066	0.109	0.069	0.112
256	32	8	0.046	0.107	0.056	0.105	0.044	0.097	0.056	0.094
256	16	16	0.047	0.094	0.058	0.115	0.037	0.085	0.064	0.108
256	8	16	0.059	0.098	0.061	0.109	0.047	0.085	0.046	0.085
512	64	8	0.057	0.096	0.041	0.084	0.041	0.088	0.049	0.094
512	32	16	0.041	0.089	0.056	0.107	0.052	0.101	0.058	0.091
512	16	32	0.046	0.084	0.046	0.098	0.057	0.095	0.048	0.087
512	8	64	0.036	0.089	0.05	0.091	0.043	0.083	0.055	0.1

level even for $T = 128$ and it can be seen that the results are basically not affected by the choice of N in these cases. For the model (5.2) we obtain very good results for $\phi \in \{-0.5, 0, 0.5\}$ while the nominal level is overestimated for $|\phi| = 0.9$ and smaller T. However, the approximation becomes much more precise if T grows and is also robust with respect to different choices of the window length N.

The results for the case $d = 0.4$ are presented in Table 3 and Table 4 and we can draw exactly the same picture from it as for $d = 0.2$. In fact, apart from the process (5.2) with $\phi = 0.9$, the performance under the null hypothesis does not change at all with different d.

5.3. Power of the test

To study the power of our test we consider the following three time varying FARIMA$((1, d, 1))$ models

$$X_{t,T} = \sqrt{\sin(\pi t/T)} Z_t^{(d)} \tag{5.4}$$

$$X_{t,T} = Z_t^{(d)} + 1.1 \cos\left(1.5 - \cos(4\pi t/T)\right) Z_{t-1}^{(d)} \tag{5.5}$$

$$\left(1 + 0.9\sqrt{t/T} B\right) X_{t,T} = Z_t^{(d)} \tag{5.6}$$

Table 3
Rejection probabilities of the test (4.10) under the null hypothesis. The data was generated according to model (5.1) with $d = 0.4$, $\theta = 0$ and different values for ϕ

			$\phi = -0.9$		$\phi = -0.5$		$\phi = 0$		$\phi = 0.5$		$\phi = 0.9$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
128	16	8	0.138	0.174	0.056	0.104	0.06	0.091	0.096	0.138	0.18	0.256
128	8	16	0.126	0.168	0.083	0.124	0.059	0.107	0.088	0.139	0.153	0.219
256	32	8	0.08	0.116	0.044	0.078	0.05	0.087	0.047	0.099	0.12	0.196
256	16	16	0.082	0.125	0.043	0.075	0.052	0.09	0.055	0.101	0.111	0.173
256	8	32	0.071	0.107	0.055	0.096	0.045	0.097	0.064	0.112	0.084	0.13
512	64	8	0.051	0.1	0.041	0.089	0.044	0.083	0.029	0.067	0.061	0.124
512	32	16	0.053	0.104	0.049	0.094	0.038	0.09	0.057	0.097	0.082	0.145
512	16	32	0.063	0.111	0.053	0.105	0.056	0.112	0.049	0.086	0.074	0.129
512	8	64	0.051	0.096	0.051	0.094	0.042	0.089	0.056	0.11	0.067	0.117

Table 4
Rejection probabilities of the test (4.10) under the null hypothesis. The data was generated according to model (5.1) with $d = 0.4$, $\phi = 0$ and different values for θ

			$\theta = -0.9$		$\theta = -0.5$		$\theta = 0.5$		$\theta = 0.9$	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
128	16	8	0.086	0.136	0.081	0.13	0.053	0.084	0.069	0.099
128	8	16	0.085	0.128	0.065	0.11	0.069	0.11	0.073	0.11
256	32	8	0.07	0.116	0.059	0.096	0.039	0.07	0.05	0.096
256	16	16	0.069	0.119	0.076	0.133	0.053	0.09	0.04	0.089
256	8	16	0.043	0.087	0.068	0.111	0.051	0.099	0.051	0.112
512	64	8	0.052	0.109	0.037	0.079	0.051	0.085	0.046	0.105
512	32	16	0.068	0.119	0.05	0.103	0.053	0.099	0.042	0.095
512	16	32	0.056	0.101	0.054	0.106	0.045	0.084	0.056	0.11
512	8	64	0.056	0.102	0.065	0.101	0.054	0.098	0.043	0.082

with $Z_t^{(d)} = (1-B)^{-d} Z_t$ for independent and standard Gaussian Z_t and different values of d. We also simulate the time varying fractional noise processes

$$X_{t,T} = (1-B)^{-d(t/T)} Z_t \tag{5.7}$$

with either $d_1(u) = 0.4u^2$ or $d_2(u) = 0.1 \times 1(u \leq 0.5) + 0.4 \times 1(u > 0.5)$. Here, in contrast to the models (5.4)–(5.6), the long memory parameter $d(u)$ varies over time. Additionally, we consider the periodic series

$$X_{t,T} = \sin(t\pi/30)(1-B)^{-d} Z_t \tag{5.8}$$

for $d = 0.2$, which in contrast to (5.4) is not locally stationary, because the function by which the innovation is multiplied depends on t instead of t/T.

The results for the alternatives (5.4)–(5.6) are depicted in Table 5, and it is remarkable that the choice of N seems to affect the results more than under the null hypothesis. This is less important for model (5.4), for which the observed rejection frequencies are large even for small sample sizes, whereas the effect can have an extreme impact on the power for the other ones; see first and foremost model (5.5) for $d = 0.2$. We display the results for the alternatives (5.7) and (5.8) in Table 6. Concerning model (5.7), it can be seen that for these kinds of processes the power seems to grow slower in T than for the alternatives (5.4)–(5.6). Again, the sensitivity of the results with respect to the choice of N is

Table 5
Rejection probabilities of the test (4.10) for the models (5.4)–(5.6)

				(5.4)		(5.5)		(5.6)	
T	N	M	d	5%	10%	5%	10%	5%	10%
128	16	8	0.2	0.694	0.811	0.198	0.303	0.028	0.084
128	8	16	0.2	0.702	0.824	0.169	0.266	0.023	0.071
256	32	8	0.2	0.909	0.968	0.211	0.332	0.132	0.262
256	16	16	0.2	0.946	0.978	0.197	0.312	0.121	0.3
256	8	16	0.2	0.942	0.98	0.158	0.264	0.164	0.32
512	64	8	0.2	0.997	1.0	0.519	0.791	0.557	0.742
512	32	16	0.2	0.999	1.0	0.477	0.702	0.575	0.764
512	16	32	0.2	1.0	1.0	0.362	0.564	0.648	0.808
512	8	64	0.2	1.0	1.0	0.258	0.39	0.664	0.823
128	16	8	0.4	0.517	0.659	0.217	0.326	0.027	0.056
128	8	16	0.4	0.649	0.769	0.188	0.262	0.022	0.067
256	32	8	0.4	0.639	0.771	0.198	0.308	0.115	0.246
256	16	16	0.4	0.795	0.903	0.162	0.292	0.11	0.271
256	8	16	0.4	0.907	0.963	0.137	0.236	0.138	0.312
512	64	8	0.4	0.731	0.861	0.275	0.525	0.471	0.652
512	32	16	0.4	0.925	0.974	0.355	0.602	0.531	0.718
512	16	32	0.4	0.989	0.995	0.355	0.564	0.662	0.784
512	8	64	0.4	0.997	1.0	0.221	0.386	0.677	0.819

Table 6
Rejection probabilities of the test (4.10) for the models (5.7) and (5.8)

			(5.7), $d_1(u)$		(5.7), $d_2(u)$		(5.8)	
T	N	M	5%	10%	5%	10%	5%	10%
128	16	8	0.058	0.108	0.037	0.075	0.349	0.552
128	8	16	0.078	0.129	0.07	0.114	0.416	0.586
256	32	8	0.054	0.108	0.049	0.125	0.147	0.239
256	16	16	0.074	0.147	0.047	0.109	0.222	0.368
256	8	16	0.094	0.143	0.085	0.128	0.288	0.426
512	64	8	0.175	0.288	0.283	0.439	0.223	0.331
512	32	16	0.131	0.218	0.218	0.356	0.276	0.378
512	16	32	0.074	0.145	0.096	0.179	0.300	0.454
512	8	64	0.104	0.172	0.099	0.181	0.332	0.467

rather large, where the best overall performance is obtained if we choose N large.

A slightly different picture can be drawn by looking at the results for the not even locally stationary model (5.8). In this case, the power decreases as T increases from 128 to 256, but gets larger as well if T grows further. However, for all sample sizes, the rejection frequencies are far above the nominal level.

5.4. Tree ring data

In this section we apply our procedure to a centered series containing 1990 annual pinus longaeva tree ring width measurements at Mammoth Creek, Utah, between 0 AD to 1989 AD. These data are displayed in the left panel of Figure 1 and were analyzed by several authors in the framework of locally stationary long-range dependent models [cf. Beran (2009) or Palma and Olea (2010) among others]. By employing the WhittleFit function from the R-package fArma to

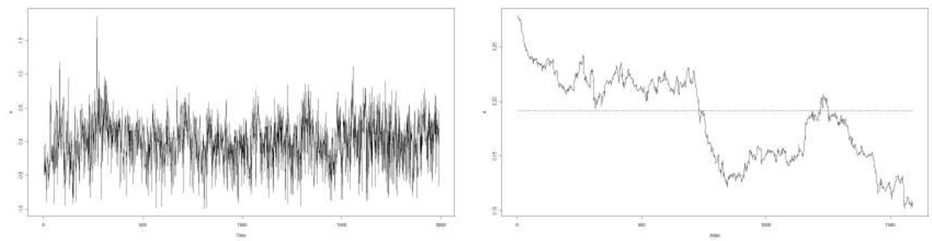

Fig 1. *Left panel: tree ring data. Right panel: the solid line represents the time varying estimator* $\hat{d}(u)$ *while the dashed horizontal line indicates the stationary estimator* $\hat{D}$.

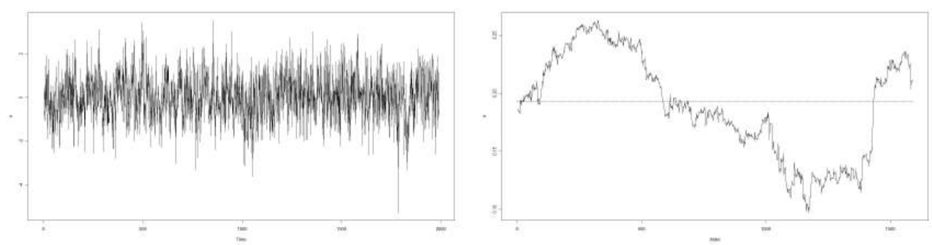

Fig 2. *Left panel: fractional noise with parameter* $d = 0.1919$. *Right panel: the solid line represents the time varying estimator* $\hat{d}(u)$ *while the dashed horizontal line indicates the stationary estimator* $\hat{D}$.

the tree ring data X_t one obtains an estimator $\hat{D} = 0.1919$ for the long memory parameter d, and if a time varying model is assumed, then the (time varying) long memory parameter $d(u)$ can be (for example) estimated through a rolling window of $N = 400$ data. Such an estimator $\hat{d}(u)$ is depicted in the right panel of Figure 1 and it suggests that $d(u)$ is not constant over time; see Beran (2009) or Palma and Olea (2010) for a similar argumentation and an approach to fit a time varying fractional noise process to the data.

However, if we simulate 1990 data from a stationary fractional noise with long memory parameter $d = 0.1919$ and calculate $\hat{d}(u)$ as above, one observes that the variability of the estimator is quite large as well; cf. Figure 2. Thus, by using graphical methods only, it is hard to tell whether the time variation of $\hat{d}(u)$ is due to non stationarity or to standard estimation errors. For this reason, we apply the test (4.10) with $N = 248$ and $N = 124$ and obtain p-values of 0.34 and 0.135, respectively. Both values do not provide enough evidence to reject the null hypothesis of stationarity at a 10% level. Even though the rejection frequencies for model (5.7) show that our test is rather conservative in models with a time varying long memory parameter, its power improves as N increases. Therefore our results indicate that the assumption of stationarity should probably not be rejected too hastily.

In order to support this observation we compare the forecasting performance of a stationary fractional noise model with that of a time varying version. For this

Table 7
The ratio $MSE^{S}(p)/MSE^{LS,N}(p)$ for different time horizons (p) and window lengths (N). A ratio smaller than one indicates that the stationary model provides a more precise forecast

Horizon p	N=100	N=125	N=150	N=175	N=200	N=500
10	0.21	0.41	0.51	0.73	0.9	0.96
20	0.10	0.375	0.33	0.66	0.87	0.94
30	0.06	0.32	0.22	0.60	0.86	0.94
40	0.04	0.28	0.16	0.57	0.85	0.93
50	0.02	0.25	0.13	0.53	0.84	0.94

reason, we start at the 100th observation and use all previous data to estimate the long memory parameter in the stationary case while we employ only the last N data in the time varying fractional noise model. We define $f_{t,r}^{S}$ as the r-step ahead forecast of the stationary model and $f_{t,r}^{LS,N}$ as the corresponding forecast for the locally stationary one. In order to compare the performance of both approaches we follow Starica and Granger (2005) and define aggregated versions through

$$\overline{X}_{t,r} = \sum_{i=1}^{r} X_{t+i}, \quad \overline{f}_{t,r}^{S} = \sum_{i=1}^{r} f_{t,i}^{S}, \quad \overline{f}_{t,r}^{LS,N} = \sum_{i=1}^{r} f_{t,i}^{LS,N}.$$

Both approaches are then compared via the ratio $MSE^{S}(r)/MSE^{LS,N}(r)$, where

$$\begin{aligned} MSE^{S}(r) &= \frac{1}{1890-r} \sum_{t=100}^{1990-r} (\overline{X}_{t,r} - \overline{f}_{t,r}^{S})^2 \\ MSE^{LS,N}(r) &= \frac{1}{1990-\max(100,N)-r} \sum_{t=\max(100,N)}^{1990-r} (\overline{X}_{t,r} - \overline{f}_{t,r}^{LS,N})^2. \end{aligned}$$

A ratio smaller than one indicates a better performance of the stationary model while a ratio bigger than one suggests that the locally stationary model provides a better forecast. The results are depicted in Table 7. From these it can be observed that the stationary model in general outperforms the time varying version. While the differences become smaller if N grows, the discrepancy is quite large for N smaller than 175. We conclude that relying on a plain visual inspection of the behaviour of the long memory parameter over time may lead to worse results in term of prediction and recommend to use a formal test for stationarity as well.

5.5. S&P 500 returns

Finally we apply the test (4.10) to 4097 observations of the S& P 500 which were recorded between April 10, 1996 and July 13, 2012. We consider the log returns $X_t = \log(Y_{t+1}/Y_t)$ $(t = 1, \dots, 4096)$ which are plotted in the right panel of Figure 3. We observe that days with either small or large movements are likely

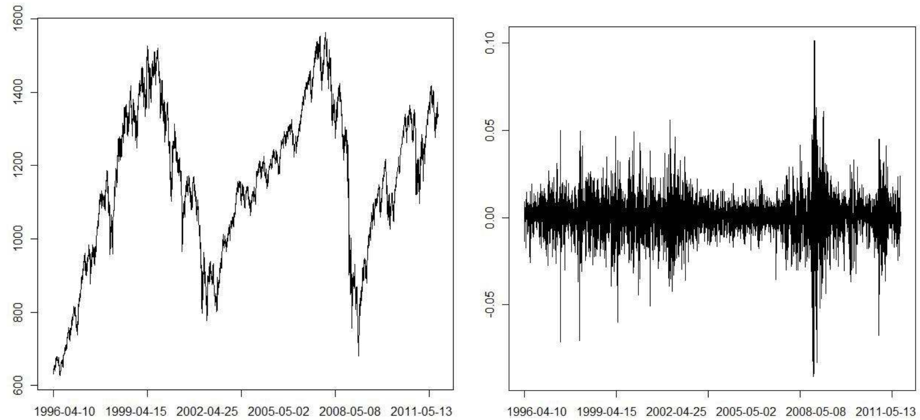

Fig 3. *The left panel displays the price of the S&P 500 between April 10th 1996 and July 13th 2012 whereas the log returns of the S&P 500 in the same period are shown in the right panel.*

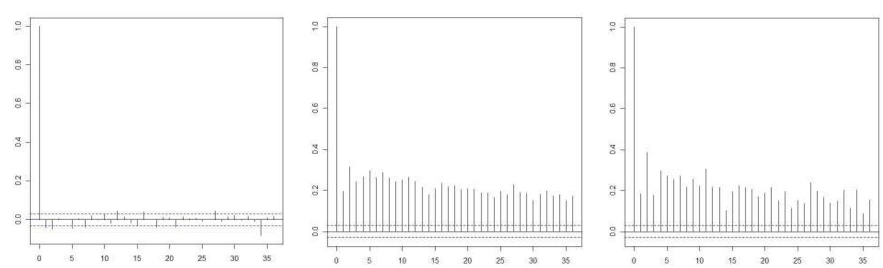

Fig 4. *Left panel: ACF (autocorrelation function) of the log returns X_t, middle panel: ACF of the absolute log returns $|X_t|$, right panel: ACF of the squared log returns X_t^2.*

to be followed by days with similar fluctuation. This effect is called 'volatility clustering' and serves as the usual motivation to employ GARCH(p, q) processes in the modelling of stock returns.

In Figure 4 the ACF (autocorrelation function) is plotted for the log returns X_t (left panel), the absolute values $|X_t|$ (middle panel) and squared returns X_t^2 (right panel). It can be seen that the autocorrelation function $\gamma(k)$ of the log returns is rather small if $k \neq 0$. However, if we take the absolute values $|X_t|$ or the squared returns X_t^2 then $\gamma(k)$ decays to zero very slow as $k \to \infty$. The latter observation is the main reason to use a long memory model if the volatility of a financial asset is analyzed.

It was shown in Mikosch and Starica (2004) and Fryzlewicz et al. (2006) that all these effects can also occur if model (1.1) is used. Starica and Granger (2005), among others, demonstrated that a simple and natural model like (1.1) is leading to a superior volatility forecast compared to a GARCH or a long range dependent FARIMA model. So it might be beneficial to consider not only

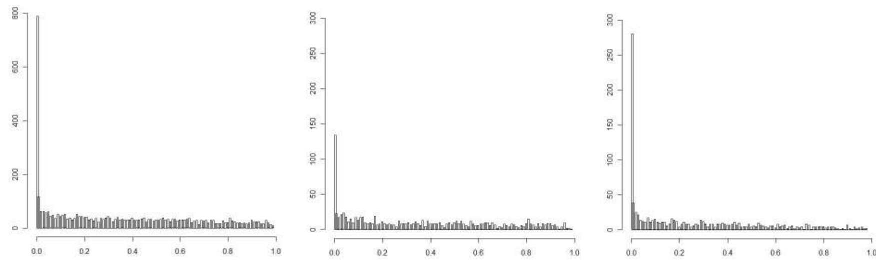

Fig 5. *The left panel displays the histogram of the p-values if the test* (4.10) *(with* $T = 64$ *and* $N = 8$*) is applied on a rolling window of the* 4096 *datapoints. In the middle panel we present the histogram of the p-values if the test* (4.10) *(with* $T = 64$ *and* $N = 8$*) is applied on a rolling window of the first* 1000 *datapoints. The right panel shows the corresponding histogram if the last* 1000 *datapoints are used.*

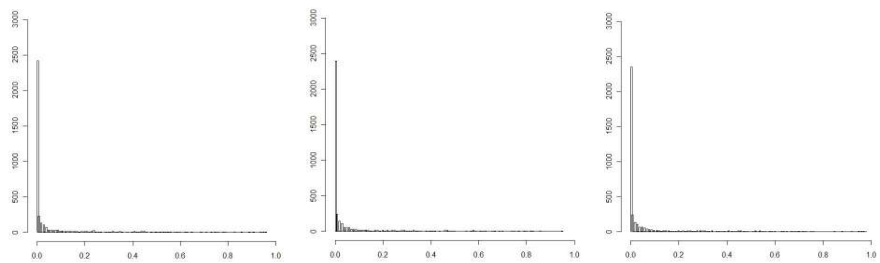

Fig 6. *Histograms of the p-values if the test* (4.10) *with* $T = 256$ *and different choice for* N *is applied on a rolling window of the* 4096 *datapoints. Left panel:* $N = 32$*, middle panel:* $N = 16$*, right panel:* $N = 8$.

complicated (e.g. long-range dependent) stationary processes in the analysis of a financial time series but to take into account models which are not stationary anymore.

We applied our test (4.10) with $T = 64$ and $N = 8$ to a rolling window of the 4096 log returns, i.e. we employed our approach using the data $X_i, \ldots X_{i+63}$ for $i = 1, \ldots, 4033$. Thus we obtain 4033 p-values whose histogram is displayed in in the left panel of Figure 5. It can be seen that the assumption of stationarity is usually not justified since for example 789 of the 4033 p-values are equal to zero and 1789 are smaller than 0.2. This effect becomes even more evident if we use a rolling window of $T = 256$ data. In this case we obtain 3841 p-values whose histograms are presented in Figure 6 for different window lengths N. If we take $N = 32$ then 2413 of the 3841 p-values are equal to zero and 3300 are smaller than 0.2. So the more data we look at, the bigger is the urgency to employ also non stationary processes in the statistical analysis. Moreover, we observe that the histograms in Figure 6 look similar and therefore the results are basically not affected by the choice of N.

One interesting observation is that during the period we took into account the data seem to become more non stationary in time which can be oberserved from the two histograms in the middle and the right panel of Figure 5. In the middle panel we display the histogram of the p-values if our test (with $T = 64$ and $N = 8$) is applied to $X_i, \dots X_{i+63}$ with $i = 1, \dots, 1000$ while the same is shown in the right panel if our approach is applied to $X_i, \dots X_{i+63}$ with $i = 3034, \dots, 4033$. If we look at both histograms it can be seen that there is a significant shift towards lower p-values.

6. Appendix: Proofs

In this section we present the proofs of all results above. We define

$$\phi_{v,\omega,T}(u,\lambda) := \Big(I_{[0,\frac{\lfloor vM\rfloor}{M}]}(u) - \lfloor vM\rfloor/M\Big) I_{[0,\frac{2\pi\lfloor\omega\frac{N}{2}\rfloor}{N}]}(\lambda)$$

for $u, \lambda \geq 0$, $v, \omega \in [0,1]$,

$$\rho_{2,T,D}(y_1,y_2) := \Big(\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}(\phi_{v_1,\omega_1,T}(u_j,\lambda_k) - \phi_{v_2,\omega_2,T}(u_j,\lambda_k))^2\frac{1}{\lambda_k^{4D}}\Big)^{1/2} \quad (6.1)$$

for $y_i = (v_i, \omega_i) \in [0,1]^2$ and set

$$\phi_{v,\omega}(u,\lambda) := \lim_{T\to\infty}\phi_{v,\omega,T}(u,\lambda) = \big(I_{[0,v]}(u) - v\big)\, I_{[0,\pi\omega]}(\lambda), \quad v,\omega \in [0,1]. \quad (6.2)$$

Note that M and N depend on T and observe the relations

$$\hat{E}_T(v,\omega) = \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k) I_N(u_j,\lambda_k)$$

$$E_T(v,\omega) = \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k) f(u_j,\lambda_k)$$

which will be employed in the proofs of the following two main theorems. All results below are assumed to hold uniformly in v, ω unless otherwise stated.

Theorem 6.1. *Suppose Assumption 2.1 holds and assume that $v, \omega, v_i, \omega_i \in [0,1]$ for $i \in \mathbb{N}$. Then we have with the notation of Theorem 3.1*

a) $\mathbb{E}(\hat{E}_T(v,\omega)) = E_T(v,\omega) + C_T(v,\omega) + O(1/T)$.

b) $\mathrm{Cov}(\hat{E}_T(v_1,\omega_1), \hat{E}_T(v_2,\omega_2))$

$$= \frac{1}{T^2}\sum_{j=1}^{M}\sum_{k=1}^{\lfloor\min(\omega_1,\omega_2)N/2\rfloor}\big(1_{[0,v_1]}(u_j) - v_1\big)\big(1_{[0,v_2]}(u_j) - v_2\big) f^2(u_j,\lambda_k)$$
$$+ O(\log(N)^2/(TN^{1-4D})) + O(N/T^2).$$

c) $\mathrm{cum}(\hat{E}_T(v_1,\omega_1), \dots, \hat{E}_T(v_l,\omega_l)) = o(T^{-l/2})$ *for $D < 1/4$ and $l \geq 3$.*

d) $\mathbb{E}|\hat{G}_T(v_1,\omega_1) - C_T(v_1,\omega_1) - (\hat{G}_T(v_2,\omega_2) - C_T(v_2,\omega_2))|^k$

$\leq (2k)! C^k \rho_{2,T,D}\left((v_1,\omega_1),(v_2,\omega_2)\right)^k T^{-k/2}$ *for all even $k \in \mathbb{N}$.*

Theorem 6.1 is the main tool for proving the results from Section 3. Regarding the bootstrap, the next theorem ensures that the random variable $\hat{E}_T^*(v,\omega)$ can be approximated by the random variable $\hat{E}_{T,a}^*(v,\omega)$ and is mainly used for proving Lemma 4.4.

Theorem 6.2. *Suppose that Assumptions 2.1, 4.1 and 4.2 are satisfied, and let* $v, \omega, v_i, \omega_i \in [0,1]$ *for* $i = 1, 2$. *Let* $\alpha > 0$ *be fixed and denote with* $A_T(\alpha)$ *the set where* $|\hat{D} - D| \leq \alpha/4$ *and*

$$|\hat{\psi}_{l,p} - \tilde{\psi}_l| l^{1-\max(\hat{D},D)} \leq C \frac{p^5 \log(T)}{\sqrt{T}} \quad \forall l \in \mathbb{N} \tag{6.3}$$

is fulfilled. Then we have

a) $\mathbb{E}\Big((\hat{E}_T^*(v,\omega) - \hat{E}_{T,a}^*(v,\omega)) 1_{A_T(\alpha)}\Big) = 0.$

b) $\operatorname{Var}\Big((\hat{E}_T^*(v,\omega) - \hat{E}_{T,a}^*(v,\omega)) 1_{A_T(\alpha)}\Big)$
$= O\Big(p^{10} \log(T)^2 \log(N)^2 N^{\max(4D-1,0)+\alpha} T^{-2}\Big).$

c) $\mathbb{E}\Big(\big|(\hat{E}_T^*(v_1,\omega_1) - \hat{E}_{T,a}^*(v_1,\omega_1)) - (\hat{E}_T^*(v_2,\omega_2) - \hat{E}_{T,a}^*(v_2,\omega_2))\big|^k 1_{A_T(\alpha)}\Big)$
$\leq (2k)! C^k \tilde{\rho}^k((v_1,\omega_1),(v_2,\omega_2)) (p^{10} \log(T)^2 N^{\max(4D-1,0)+\alpha} T^{-2})^{k/2}$
for all $k \in \mathbb{N}$ *even, where* $\tilde{\rho}((v_1,\omega_1),(v_2,\omega_2)) := 1_{\{v_1 \neq v_2 \text{ or } \omega_1 \neq \omega_2\}}$.

We begin with the proof of Theorem 6.1 and 6.2 for which we require some technical lemmata.

Lemma 6.3. *Suppose Assumption 2.1 is satisfied. Then for all* $\lambda \in (0, \pi)$ *and* $N \in \mathbb{N}$

$$\Big| \sum_{\substack{l,m=0 \\ |l-m|>N}}^{\infty} \psi_l(u) \psi_m(u) \exp(-i\lambda(l-m)) \Big| \leq \frac{C}{\lambda N^{1-2D}}.$$

Proof. Without loss of generality we only consider the case $m > l$. We have

$$\sum_{\substack{l,m=0 \\ m-l>N}}^{\infty} \psi_l(u) \psi_m(u) \exp(-i\lambda(l-m)) = \sum_{l=0}^{\infty} \psi_l(u) \sum_{m=l+N+1}^{\infty} \psi_m(u) \exp(-i\lambda(l-m)),$$

and the absolute value of the right term can be bounded through

$$\sum_{l=0}^{\infty} \Big| \psi_l(u) \exp(-i\lambda l) \Big| \Bigg(\Big| \sum_{m=l+N+1}^{\infty} \frac{a(u)}{m^{1-d(u)}} \exp(i\lambda m) \Big| + \sum_{m=l+N+1}^{\infty} \Big| \psi_m(u) - \frac{a(u)}{m^{1-d(u)}} \Big| \Bigg) \tag{6.4}$$

where $a(u)$ is the function from (2.6). Equation (2.9) in chapter 5 of Zygmund (1959) says that

$$\Big|\sum_{m=l+N+1}^{\infty} \frac{1}{m^{1-d(u)}} \exp(-i\lambda m)\Big| \leq \frac{C}{\lambda} \frac{1}{(l+N)^{1-D}}$$

holds for a constant $C \in \mathbb{R}$ which is independent of l, N, u and λ. In addition, (2.6) implies

$$\sup_u |\psi_l(u)| \leq C|l|^{D-1} \quad \forall l \geq 1. \tag{6.5}$$

If we combine the last two statements with (2.6) we can bound (6.4) up to a constant through

$$\sum_{l=1}^{\infty} \frac{1}{l^{1-D}} \left(\frac{1}{\lambda} \frac{1}{(l+N)^{1-D}} + \frac{1}{(l+N)^{1-D}} \right) \leq \frac{C}{\lambda} \frac{1}{N^{1-2D}}. \qquad \square$$

Lemma 6.4. *a) For all $n \geq 1$ and $k_1, k_2 \in \mathbb{N}$ there exists a constant $C(k_1, k_2) > 0$ such that:*

$$\sum_{\substack{l,m=1 \\ |l-m| \geq n}}^{\infty} \frac{\log^{k_1}|l| \log^{k_2}|m|}{|lm|^{1-D}} \frac{1}{|l-m|} \leq C(k_1,k_2) \left(\frac{\log^{k_1+k_2+1}(n)}{n^{1-2D}} + 1_{\{n=1\}} \right).$$

b) For $n \geq 1$ we have

$$\sum_{\substack{l,m=1 \\ 0<|l-m|<n}}^{\infty} \frac{1}{|lm|^{1-D}} \leq C n^{2D}.$$

c) We write $(+)_{\neq}$ if $|m_1 - l_2| \leq n$, $|m_2 - l_1| \leq n$ and $m_1 - l_2 + m_1 - l_1 \neq 0$ hold. Then for $n \geq 2$

$$\frac{1}{n} \sum_{\substack{m_1,m_2,l_1,l_2=1 \\ (+)_{\neq}}}^{\infty} \frac{1}{|m_1 m_2 l_1 l_2|^{1-D}} \frac{|m_2 - l_1|}{|m_1 - l_2 + m_2 - l_1|} \leq \frac{C \log(n)}{n^{1-4D}}.$$

d) For $l \geq 3$ we write $(+)_{\neq,l}$ if $|n_1 - m_l| \leq n$ and $|n_{i+1} - m_i| \leq n$ are satisfied for $i \in \{1, \ldots, l-1\}$ and furthermore $m_1 - n_1 + m_2 - n_2 + \cdots + m_l - n_l \neq 0$ holds. Then there exists $C_l > 0$ such that for all $n \geq 2$

$$\sum_{\substack{m_i,n_i=1 \\ (+)_{\neq,l}}}^{\infty} \frac{1}{|m_1 n_1 m_2 n_2 \cdots m_l n_l|^{1-D}} \frac{1}{|m_1 - n_1 + m_2 - n_2 + \cdots + m_l - n_l|} \leq C_l \log(n) n^{2Dl-4D}.$$

Proof. Before we begin with the proof, note that a simple change of variables yields

$$\int_a^b \frac{1}{x^{1-D}} \frac{\log^k x}{(c \pm x)^e} dx = \frac{1}{c^{e-D}} \int_{a/c}^{b/c} \frac{\log^k(cz)}{z^{1-D}} \frac{1}{(1 \pm z)^e} dz \tag{6.6}$$

for $a, b, c, e \in I\!R$ with $a \leq b$, $c > 0$, $k \in \mathbb{N}$ if any of the integrals exist. The proof now basically works by considering approximating integrals instead of the sums, using (6.6) and afterwards employing that

$$\int_a^b \frac{|\log^k(z)|}{z^{1-D}} \frac{1}{1-z} dz \leq C(k) + C|\log(1-b)| \tag{6.7}$$

holds for $k \in \mathbb{N}_0$, $0 < a < b < 1$ and constants $C(k) \in I\!R$ which are independent of a and b. Note that the absolute value of the right hand side of (6.6) is bounded by

$$\frac{1}{c^{e-D}} \int_0^\infty \frac{|\log^k(cz)|}{z^{1-D}} \frac{1}{(1 \pm z)^e} dz$$

which is in any case finite if $0 < D < 1$, $0 < e < 1$ and $1-D+e > 1$. If $e = 1$ and b/c is close to the possible pole 1, (6.7) implies that the integral on the right hand side of (6.6) is only bounded by a constant times some additional log term which incorporates in some way how close the boundary is to 1. This rule of thumb will be helpful in understanding the treatment of the approximating integrals in the following. Since all proofs work in that particular way of replacing the sum through integrals and applying (6.6) and (6.7) afterwards we will present the details for part a) only.

Proof of a): For $n = 1$ let $l > m$ without loss of generality. Then

$$\sum_{m=1}^\infty \sum_{l=m+1}^\infty \frac{\log^{k_1}|l| \log^{k_2}|m|}{|lm|^{1-D}} \frac{1}{|m-l|} = \sum_{m=1}^\infty \sum_{s=1}^\infty \frac{\log^{k_2}(m)}{m^{1-D}(m+s)^{1-D-\varepsilon}} \frac{\log^{k_1}(m+s)}{s(m+s)^\varepsilon},$$

for some ε small enough such that $2 - 2D - \varepsilon > 1$. Both sums are finite then, so let $n \geq 2$. Again, we discuss the case $l > m$ only, which becomes

$$\sum_{l=n+1}^\infty \frac{\log^{k_1} l}{l^{1-D}} \sum_{m=2}^{l-n} \frac{\log^{k_2} m}{m^{1-D}} \frac{1}{l-m}.$$

If we treat the expression in the second summand as a function in m, it can be seen that this function only has a finite number of points where the first derivate equals zero. Thus it is piecewise monotonic, which allows us to bound the sum over m by its approximating integral, i.e. by

$$\int_1^{l-n+1} \frac{\log^{k_2} x}{x^{1-D}} \frac{1}{l-x} dx = \frac{1}{l^{1-D}} \int_{1/l}^{1-\frac{n-1}{l}} \frac{\log^{k_2}(lz)}{z^{1-D}} \frac{1}{1-z} dz$$

$$\leq \frac{\log^{k_2}(l)}{l^{1-D}} \int_{1/l}^{1-\frac{n-1}{l}} \frac{1}{z^{1-D}} \frac{1}{1-z} dz.$$

With (6.7) it follows that the entire expression can be (up to a further constant) bounded by

$$\sum_{l=n+1}^{\infty} \frac{\log^{k_1+k_2} l}{l^{2-2D}} \left(1 + \left|\log\left(\frac{n-1}{l}\right)\right|\right) \leq 3 \sum_{l=n+1}^{\infty} \frac{\log^{k_1+k_2+1} l}{l^{2-2D}} = O\left(\frac{\log^{k_1+k_2+1} n}{n^{1-2D}}\right).$$

This yields the claim for $m > 0$ and we now consider the case $m < 0$. A straightforward calculation yields that

$$\sum_{l \geq n/2, m \leq \min(0, l-n)}^{\infty} \frac{\log^{k_1}|l| \log^{k_2}|m|}{|lm|^{1-D}} \frac{1}{|m-l|}$$
$$\leq \sum_{l=n/2}^{n-1} \frac{\log^{k_1} l}{l^{1-D}} \left(\frac{\log^{k_2} n}{n} + \sum_{m=n-l+1}^{\infty} \frac{\log^{k_2} m}{m^{1-D}} \frac{1}{l+m}\right)$$
$$+ \sum_{l=n}^{\infty} \frac{\log^{k_1} l}{l^{1-D}} \sum_{m=2}^{\infty} \frac{\log^{k_2} m}{m^{1-D}} \frac{1}{l+m},$$

and by replacing the sum over m through its approximating integral we can bound this expression by

$$\frac{\log^{k_2} n}{n} \sum_{l=n/2}^{n-1} \frac{\log^{k_1} l}{l^{1-D}} + \sum_{l=n/2}^{n-1} \frac{\log^{k_1} l}{l^{1-D}} \int_{n-l}^{\infty} \frac{\log^{k_2} x}{x^{1-D}} \frac{1}{l+x} dx$$
$$+ \sum_{l=n}^{\infty} \frac{\log^{k_1} l}{l^{1-D}} \int_{1}^{\infty} \frac{\log^{k_2} x}{x^{1-D}} \frac{1}{l+x} dx.$$

By using (6.6) we can bound both integrals through a constant times $\log^{k_2}(l)/l^{1-D}$ which then yields the claim by calculating the resulting sums. □

Analogously to the above proof we can show the next lemma, which, although it looks similar to Lemma 6.4 (and is proven in the same way), is different since the index of summation m is fixed.

Lemma 6.5. *For all $m \in \mathbb{Z}$ and $n \geq 1$ we have*

a) $$\sum_{\substack{l=1 \\ 0<|l-m|<n}}^{\infty} \frac{1}{|l|^{1-D}} \frac{1}{|l-m|} \leq C\left(\frac{\log|m|}{|m|^{1-D}} 1_{\{m \neq 0\}} + 1_{\{m=0\}}\right) \leq C$$

b) $$\sum_{\substack{l=1 \\ n/2 \leq |l-m| < n}}^{\infty} \frac{1}{|l|^{1-d}} \frac{1}{n-|l-m|}$$
$$\leq C\left(\max\left(\frac{\log|n-m|}{|n-m|^{1-d}}, \frac{\log|n+m|}{|n+m|^{1-d}}\right) 1_{\{m \neq n\}} + 1_{\{m=n\}}\right) \leq C.$$

6.1. Proof of Theorem 6.1

Proof of a). We have

$$\mathbb{E}\Big(\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)I_N(u_j,\lambda_k)\Big)$$
$$= \frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{2\pi N}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}$$
$$\psi_{t_j-N/2+1+p,T,l}\psi_{t_j-N/2+1+q,T,m}\mathbb{E}(Z_{t_j-N/2+1+p-m}Z_{t_j-N/2+1+q-l})$$
$$\exp(-i\lambda_k(p-q)).$$

Set $e_{j,N} := t_j - N/2 + 1$. By using the independence of the innovations Z_i we obtain that the above term equals

$$\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{2\pi N}\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\sum_{\substack{q=0\\0\le q+m-l\le N-1}}^{N-1}\psi_{e_{j,N}+q+m-l,T,l}\psi_{e_{j,N}+q,T,m}$$
$$\exp(-i\lambda_k(m-l)). \tag{6.8}$$

Write the product of the ψ-terms above as

$$\psi_l\Big(\frac{e_{j,N}+q+m-l}{T}\Big)\psi_m\Big(\frac{e_{j,N}+q}{T}\Big)$$
$$+\psi_{e_{j,N}+q+m-l,T,l}\Big(\psi_{e_{j,N}+q,T,m}-\psi_m\Big(\frac{e_{j,N}+q}{T}\Big)\Big)$$
$$\left.+\psi_m\left(\frac{e_{j,N}+q}{T}\right)\left(\psi_{e_{j,N}+q+m-l,T,l}-\psi_l\left(\frac{e_{j,N}+q+m-l}{T}\right)\right)\right\}, \tag{6.9}$$

so (6.8) splits into a sum of three terms. We will now demonstrate that the second summand is of order $O(1/T)$ and analogously for the third one. The absolute value of the second summand can be bounded by

$$\frac{1}{M}\sum_{j=1}^{M}\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\frac{1}{2\pi N}\sum_{\substack{q=0\\0\le q+m-l\le N-1}}^{N-1}|\psi_{e_{j,N}+q+m-l,T,l}|\left|\psi_{e_{j,N}+q,T,m}-\psi_m\left(\frac{e_{j,N}+q}{T}\right)\right|$$
$$\times\left|\frac{1}{N}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\exp(-i\lambda_k(m-l))\right|. \tag{6.10}$$

We employ (A.2) of Eichler (2008) which says that there exists a constant $C \in \mathbb{R}$ such that for all $\{r \in \mathbb{Z} : r \operatorname{mod} N/2 \neq 0\}$ we have

$$\Big|\frac{1}{N}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u,\lambda_k)\exp(-i\lambda_k r)\Big| \le \frac{C}{|r \operatorname{mod} N/2|} \tag{6.11}$$

uniformly in v, ω. Using (2.5), (2.9), (6.11) and a symmetry argument we can bound (6.10) up to a constant by

$$
\begin{aligned}
&2 \sum_{\substack{l,m=1 \\ 0<|l-m|<N/2}}^{\infty} \frac{1}{|l|^{1-D}} \sup_{q,t_j} \Big| \psi_{t_j-N/2+1+q,T,m} - \psi_m\Big(\frac{t_j-N/2+1+q}{T}\Big) \Big| \frac{1}{|l-m|} \\
&+ \sum_{\substack{l,m=1 \\ |l-m|=N/2 \,\vee\, l=m}}^{\infty} \frac{1}{|l|^{1-D}} \sup_{q,t_j} \Big| \psi_{t_j-N/2+1+q,T,m} - \psi_m\Big(\frac{t_j-N/2+1+q}{T}\Big) \Big| \\
&+ \sum_{m=1}^{N/2-1} \sup_{q,t_j} \Big| \psi_{t_j-N/2+1+q,T,m} - \psi_m\Big(\frac{t_j-N/2+1+q}{T}\Big) \Big| \frac{1}{m} \\
&+ \frac{1}{T} \sum_{l=1}^{N/2-1} \frac{1}{|l|^{1-D}} \frac{1}{l} + \frac{1}{T}.
\end{aligned}
$$

Note that the terms in the final line correspond to the case where $l = 0$ or $m = 0$. It can be shown that each of the terms above is of order $O(1/T)$ due to Lemma 6.5, (2.5) and $D < 1/2$. In the following we will bound expressions like the above one w.l.o.g. by a constant times the first summand, i.e. from now on we will only consider the case $0 < |l-m| < N/2$ if we derive the order of error terms. We do this since the remaining terms will be either of the same or of smaller order and are treated analogously.

Following the above argumentation we obtain that (6.9) equals

$$
\begin{aligned}
&\frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{N/2} \phi_{v,\omega,T}(u_j,\lambda_k) \frac{1}{2\pi N} \sum_{p,q=0}^{N-1} \sum_{l,m=0}^{\infty} \psi_l\Big(\frac{e_{j,N}+p}{T}\Big) \psi_m\Big(\frac{e_{j,N}+q}{T}\Big) \\
&\quad \times \mathbb{E}(Z_{e_{j,N}+p-m} Z_{e_{j,N}+q-l}) \exp(-i\lambda_k(p-q)) + O(1/T) \\
&= \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{N/2} \phi_{v,\omega,T}(u_j,\lambda_k) \frac{1}{2\pi N} \sum_{p,q=0}^{N-1} \sum_{l,m=0}^{\infty} \mathbb{E}(Z_{e_{j,N}+p-m} Z_{e_{j,N}+q-l}) \psi_l(u_j) \psi_m(u_j) \\
&\quad \times \exp(-i\lambda_k(p-q)) + \tilde{C}_T(v,\omega) + O(1/T)
\end{aligned} \tag{6.12}
$$

with

$$
\begin{aligned}
&\tilde{C}_T(v,\omega) \\
&:= \frac{1}{T} \sum_{j=1}^{M} \sum_{k=1}^{N/2} \phi_{v,\omega,T}(u_j,\lambda_k) \frac{1}{2\pi N} \sum_{p,q=0}^{N-1} \sum_{l,m=0}^{\infty} \mathbb{E}(Z_{e_{j,N}+p-m} Z_{e_{j,N}+q-l}) \\
&\quad \times \exp(-i\lambda_k(p-q)) \Bigg\{ \left(\psi_l\left(\frac{e_{j,N}+p}{T}\right) - \psi_l(u_j) \right) \psi_m(u_j) \\
&\qquad + \left(\psi_m\left(\frac{e_{j,N}+q}{T}\right) - \psi_m(u_j) \right) \psi_l(u_j)
\end{aligned}
$$

$$+\left(\psi_l\left(\frac{e_{j,N}+p}{T}\right)-\psi_l(u_j)\right)\left(\psi_m\left(\frac{e_{j,N}+q}{T}\right)-\psi_m(u_j)\right)\Bigg\}. \qquad (6.13)$$

Let us begin with the first summand of (6.12). This term can be rewritten as

$$\begin{aligned}
&\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{2\pi N}\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\sum_{\substack{q=0\\0\le q+m-l\le N-1}}^{N-1}\psi_l(u_j)\psi_m(u_j)\\
&\times\exp(-i\lambda_k(m-l))\\
&=\frac{1}{2\pi T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\psi_l(u_j)\psi_m(u_j)\exp(-i\lambda_k(m-l))\\
&-\frac{1}{2\pi TN}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}|l-m|\psi_l(u_j)\psi_m(u_j)\exp(-i\lambda_k(m-l))\\
&=A_T(v,\omega)-B_T(v,\omega) \qquad (6.14)
\end{aligned}$$

where A_T and B_T are defined implicitly. (6.5) and (6.11) prove that B_T is up to a constant independent of (v,ω) bounded by

$$\frac{1}{N}\sum_{\substack{l,m=1\\0<|l-m|<N/2}}^{\infty}\frac{1}{l^{1-D}}\frac{1}{m^{1-D}}$$

which is of order $O(\log(N)/N^{1-2D})$ due to Lemma 6.4 b). Note that the cases with either $l=0$, $m=0$ or $N/2\le|l-m|<N$ are of the same or of smaller order. Consider A_T next. Our aim is to skip the condition $|l-m|\le N-1$. By employing Lemma 6.3 we obtain

$$\begin{aligned}
\tilde{A}_T(v,\omega)&=\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{2\pi}\sum_{\substack{l,m=0\\|l-m|\ge N}}^{\infty}\psi_l(u_j)\psi_m(u_j)\exp(-i\lambda_k(m-l))\\
&=O\left(\frac{\log(N)}{N^{1-2D}}\right),
\end{aligned}$$

uniformly in (v,ω), and therefore A_T can be decomposed as

$$\begin{aligned}
&\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)f(u_j,\lambda_k)+O\left(\frac{\log(N)}{N^{1-2D}}\right)-\tilde{A}_T(v,\omega)\\
&=E_T(v,\omega)+O\left(\frac{\log(N)}{N^{1-2D}}\right).
\end{aligned}$$

Setting $C_T=\tilde{C}_T-\tilde{A}_T-B_T$ note that all three terms are zero if the ψ_l are independent of time. For the first summand this follows immediately as it is built on differences of ψ variables, whereas for the latter two ones the claim follows by definition of $\phi_{v,\omega,T}(u,\lambda)$. We are therefore left to show that $\tilde{C}_T = O(N^2/T^2) + O(\log(N)/N^{1-2D})$ holds uniformly in $v,\omega \in [0,1]$. Without loss of generality we only consider the first summand in (6.13) which equals

$$\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{2\pi N}\sum_{\substack{p=0\\0\le p+l-m\le N-1}}^{N-1}\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\psi_l'(u_j)\psi_m(u_j)\frac{p-N/2+1}{T}$$
$$\times\exp(-i\lambda_k(m-l)) + O(N^2/T^2)$$

due to a second order Taylor expansion plus Assumption 2.1, Lemma 6.4 a) and (6.11). We proceed here as for A_T and B_T in (6.14) above, and a similar argument as for B_T proves that we can skip the condition $0 \le p+l-m \le N-1$ at the cost of an error of order $O(\log(N)/N^{1-2D})$. Therefore the above expression equals

$$\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{2\pi N}\sum_{p=0}^{N-1}\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\psi_l'(u_j)\psi_m(u_j)\frac{p-N/2+1}{T}$$
$$\times\exp(-i\lambda_k(m-l)) + O\left(\frac{\log(N)}{N^{1-2d}}\right).$$

Using $\sum_{p=0}^{N-1}(p-N/2+1)/T = N/(2T)$ we see that this term is the same as

$$\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{4\pi T}\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\psi_l'(u_j)\psi_m(u_j)\exp(-i\lambda_k(m-l))$$
$$+O\left(\frac{\log(N)}{N^{1-2d}}\right),$$

and its first part is some $O(1/T)$ because of (2.8), (6.5), (6.11) and Lemma 6.4 a) with $n=1$, $k_1=1$ and $k_2=0$. □

Proof of b). We set

$$V_T^{true} = \mathrm{Cov}\Big(\frac{1}{T}\sum_{j_1=1}^{M}\sum_{k_1=1}^{N/2}\phi_{v_1,\omega_1,T}(u_{j_1},\lambda_{k_1})I_N(u_{j_1},\lambda_{k_1}),$$
$$\frac{1}{T}\sum_{j_2=1}^{M}\sum_{k_2=1}^{N/2}\phi_{v_2,\omega_2,T}(u_{j_2},\lambda_{k_2})I_N(u_{j_2},\lambda_{k_2})\Big)$$
$$= \frac{1}{T^2}\sum_{j_1,j_2=1}^{M}\sum_{k_1,k_2=1}^{N/2}\phi_{v_1,\omega_1,T}(u_{j_1},\lambda_{k_1})\phi_{v_2,\omega_2,T}(u_{j_2},\lambda_{k_2})$$

$$\times \frac{1}{(2\pi N)^2} \sum_{p_1,p_2,q_1,q_2=0}^{N-1} \sum_{m_1,m_2,l_1,l_2=0}^{\infty} \psi_{e_{j_1,N}+p_1,T,m_1} \psi_{e_{j_1,N}+q_1,T,l_1}$$
$$\times \psi_{e_{j_2,N}+p_2,T,m_2} \psi_{e_{j_2,N}+q_2,T,l_2}$$
$$\times \ \mathrm{cum}(Z_{e_{j_1,N}+p_1-m_1} Z_{e_{j_1,N}+q_1-l_1}, Z_{e_{j_2,N}+p_2-m_2} Z_{e_{j_2,N}+q_2-l_2})$$
$$\times \exp(-i\lambda_{k_1}(p_1-q_1)) \exp(-i\lambda_{k_2}(p_2-q_2)).$$

with $e_{j_i,N} = t_{j_i} - N/2 + 1$. We start by considering the approximating version V_T^{appr} which is the same as above, but where all ψ-terms have been replaced, so e.g. $\psi_{e_{j_1,N}+p_1,T,m_1}$ by $\psi_{m_1}(u_{j_1})$ and similarly for the others. Using the well-known formula

$$\begin{aligned} &\mathrm{cum}(Z_{e_{j_1,N}+p_1-m_1} Z_{e_{j_1,N}+q_1-l_1}, Z_{e_{j_2,N}+p_2-m_2} Z_{e_{j_2,N}+q_2-l_2}) \\ &= \mathrm{cum}(Z_{e_{j_1,N}+p_1-m_1} Z_{e_{j_2,N}+q_2-l_2}) \mathrm{cum}(Z_{e_{j_1,N}+q_1-l_1} Z_{e_{j_2,N}+p_2-m_2}) \\ &+ \mathrm{cum}(Z_{e_{j_1,N}+p_1-m_1} Z_{e_{j_2,N}+p_2-m_2}) \mathrm{cum}(Z_{e_{j_1,N}+q_1-l_1} Z_{e_{j_2,N}+q_2-l_2}). \end{aligned} \tag{6.15}$$

the computation of V_T^{appr} splits into two similar terms which we denote with $V_{T,1}$ and $V_{T,2}$. We start by considering the first one. Because of the independence of the innovations Z_i we obtain

$$V_{T,1} = \frac{1}{T^2} \sum_{j_1,j_2=1}^{M} \sum_{k_1,k_2=1}^{N/2} \phi_{v_1,\omega_1,T}(u_{j_1}, \lambda_{k_1}) \phi_{v_2,\omega_2,T}(u_{j_2}, \lambda_{k_2})$$
$$\times \frac{1}{(2\pi N)^2} \sum_{m_1,m_2,l_1,l_2=0}^{\infty} \sum_{\substack{q_1,q_2=0 \\ 0 \le q_2+m_1-l_2+t_{j_2}-t_{j_1} \le N-1 \\ 0 \le q_1+m_2-l_1+t_{j_1}-t_{j_2} \le N-1}}^{N-1} \psi_{m_1}(u_{j_1}) \psi_{l_1}(u_{j_1}) \psi_{m_2}(u_{j_2}) \psi_{l_2}(u_{j_2})$$
$$\times \exp(-i\lambda_{k_1}(q_2 - q_1 + t_{j_2} - t_{j_1} + m_1 - l_2))$$
$$\times \exp(i\lambda_{k_2}(q_2 - q_1 + t_{j_2} - t_{j_1} + l_1 - m_2)).$$

We divide the sum over j_1, j_2 into two sums, namely one sum where $j_1 = j_2$ is satisfied and one sum with $j_1 \neq j_2$. Then

$$\begin{aligned} V_{T,1} = &\frac{1}{T^2} \sum_{j_1=1}^{M} \sum_{k_1,k_2=1}^{N/2} \phi_{v_1,\omega_1,T}(u_{j_1}, \lambda_{k_1}) \phi_{v_2,\omega_2,T}(u_{j_1}, \lambda_{k_2}) \\ &\times \frac{1}{(2\pi N)^2} \sum_{m_1,m_2,l_1,l_2=0}^{\infty} \sum_{\substack{q_1,q_2=0 \\ 0 \le q_2+m_1-l_2 \le N-1 \\ 0 \le q_1+m_2-l_1 \le N-1}}^{N-1} \psi_{m_1}(u_{j_1}) \psi_{l_1}(u_{j_1}) \psi_{m_2}(u_{j_1}) \psi_{l_2}(u_{j_1}) \\ &\times \exp(-i\lambda_{k_1}(q_2 - q_1 + m_1 - l_2)) \exp(i\lambda_{k_2}(q_2 - q_1 + l_1 - m_2)) + V_{T,1}^{j_1 \neq j_2}, \end{aligned} \tag{6.16}$$

where $V_{T,1}^{j_1 \neq j_2}$ corresponds to the case where j_1 and j_2 are not equal to each other. The first claim will be

$$V_{T,1}^{j_1 \neq j_2} = O\left(\log(N)^2 / (TN^{1-4D})\right), \tag{6.17}$$

i.e. the second summand in (6.16) vanishes asymptotically, meaning that we can restrict ourselves to the case $j_1 = j_2$ thereafter.

Proof of (6.17): Note that we can bound the absolute value of $V_{T,1}^{j_1\neq j_2}$ by the sum of four terms $V_{T,1,i}^{j_1\neq j_2}$ $[i = 1,\dots,4]$ which are the absolute values of the terms corresponding to the following four cases:

$$
\begin{aligned}
&1)\ q_2 - q_1 + t_{j_2} - t_{j_1} + m_1 - l_2 \neq 0 \text{ and } q_2 - q_1 + t_{j_2} - t_{j_1} + l_1 - m_2 = 0\\
&2)\ q_2 - q_1 + t_{j_2} - t_{j_1} + m_1 - l_2 = 0 \text{ and } q_2 - q_1 + t_{j_2} - t_{j_1} + l_1 - m_2 \neq 0\\
&3)\ q_2 - q_1 + t_{j_2} - t_{j_1} + m_1 - l_2 \neq 0 \text{ and } q_2 - q_1 + t_{j_2} - t_{j_1} + l_1 - m_2 \neq 0\\
&4)\ q_2 - q_1 + t_{j_2} - t_{j_1} + m_1 - l_2 = 0 \text{ and } q_2 - q_1 + t_{j_2} - t_{j_1} + l_1 - m_2 = 0
\end{aligned}
\tag{6.18}
$$

We will present the details for the term $V_{T,1,3}^{j_1\neq j_2}$ only since it is the dominating one due to the least restrictive conditions. Setting $\Delta t = t_{j_2} - t_{j_1}$ we obtain that $|V_{T,1,3}^{j_1\neq j_2}|$ equals

$$
\begin{aligned}
&\Big| \frac{1}{T^2} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{M} \sum_{k_1,k_2=1}^{N/2} \phi_{v_1,\omega_1,T}(u_{j_1},\lambda_{k_1}) \phi_{v_2,\omega_2,T}(u_{j_2},\lambda_{k_2})\\
&\times \frac{1}{(2\pi N)^2} \sum_{m_1,m_2,l_1,l_2=0}^{\infty} \sum_{\substack{q_1,q_2=0\\ 0\le q_2+m_1-l_2+\Delta t\le N-1\\ 0\le q_1+m_2-l_1-\Delta t\le N-1\\ (6.18)}}^{N-1} \psi_{m_1}(u_{j_1})\psi_{l_1}(u_{j_1})\psi_{m_2}(u_{j_2})\psi_{l_2}(u_{j_2})\\
&\times \exp(-i\lambda_{k_1}(q_2 - q_1 + \Delta t + m_1 - l_2)) \exp(-i\lambda_{k_2}(q_1 - q_2 - \Delta t + m_2 - l_1)) \Big|\\
&\le \frac{1}{(2\pi T)^2} \sum_{j_1=1}^{M} \sum_{m_1,m_2,l_1,l_2=0}^{\infty} \sum_{q_2=0}^{N-1} \sum_{\substack{j_2\neq j_1\\ 0\le q_2+m_1-l_2+\Delta t\le N-1}}^{M}\\
&\times |\psi_{m_1}(u_{j_1})\psi_{l_1}(u_{j_1})\psi_{m_2}(u_{j_2})\psi_{l_2}(u_{j_2})|\\
&\times \sum_{\substack{q_1=0\\ 0\le q_1+m_2-l_1-\Delta t\le N-1\\ (6.18)}}^{N-1} \Big| \frac{1}{N} \sum_{k_1=1}^{N/2} \phi_{v_1,\omega_1,T}(u_{j_1},\lambda_{k_1})\\
&\qquad\qquad \exp(-i\lambda_{k_1}(q_2 - q_1 + \Delta t + m_1 - l_2)) \Big|\\
&\times \Big| \frac{1}{N} \sum_{k_2=1}^{N/2} \phi_{v_2,\omega_2,T}(u_{j_2},\lambda_{k_2}) \exp(-i\lambda_{k_2}(q_1 - q_2 - \Delta t + m_2 - l_1)) \Big|.
\end{aligned}
\tag{6.19}
$$

The conditions $0 \le q_2 + m_1 - l_2 + \Delta t \le N - 1$ and $0 \le q_1 + m_2 - l_1 - \Delta t \le N - 1$ can only be satisfied if $|m_1 - l_2 + \Delta t| < N$ and $|m_2 - l_1 - \Delta t| < N$ hold. By combining this with (6.5) and (6.11) it can be seen that the above term is up to a constant bounded by

$$\frac{1}{T^2}\sum_{\substack{m_1,m_2,l_1,l_2=1\\ m_1-l_2+m_2-l_1\neq 0}}^{\infty}\sum_{q_2=0}^{N-1}\sum_{\substack{j_2\neq j_1\\ 0\leq q_2+m_1-l_2+\Delta t\leq N-1\\ |m_1-l_2+\Delta t|<N\\ |m_2-l_1-\Delta t|<N}}^{M}\frac{1}{|m_1m_2l_1l_2|^{1-D}}$$
$$\times\sum_{\substack{q_1\in A_N\\ |q_2-q_1+\Delta t+m_1-l_2|<N/2\\ |q_1-q_2-\Delta t+m_2-l_1|<N/2}}\frac{1}{|q_2-q_1+\Delta t+m_1-l_2|}\frac{1}{|q_1-q_2-\Delta t+m_2-l_1|} \tag{6.20}$$

where $A_N = \{0,1,2,\ldots,N-1\}\backslash\{z_1,z_2\}$ with $z_1 = q_2+\Delta t+m_1-l_2$, $z_2 = q_2+\Delta t+l_1-m_2$. We used once more that the cases with $m_i = 0$, $l_i = 0$, $m_1-l_2+m_2-l_1 = 0$, $|q_2-q_1+\Delta t+m_1-l_2| \geq N/2$ or $|q_1-q_2-\Delta t+m_2-l_1| \geq N/2$ are of the same or smaller order and that z_1 and z_2 correspond to the values of q_1 for which the argument in one of the exp-function is zero which cannot occur because of (6.18). By considering the approximating integral we can bound the latter sum up to a constant by

$$\int_A \frac{1}{|q_2-q_1+\Delta t+m_1-l_2|}\frac{1}{|q_1-q_2-\Delta t+m_2-l_1|}dq_1$$

with $A = [0,N-1]\backslash\{[z_1-1,z_1+1]\cup[z_2-1,z_2+1]\}$. A simple integration via a decomposition into partial fractions yields that (6.20) is thus (up to a constant) bounded by

$$\frac{1}{T^2}\sum_{j_1=1}^{M}\sum_{\substack{m_1,m_2,l_1,l_2=1\\ m_1-l_2+m_2-l_1\neq 0}}^{\infty}\sum_{q_2=0}^{N-1}\sum_{\substack{j_2\neq j_1\\ 0\leq q_2+m_1-l_2+\Delta t\leq N-1\\ |m_1-l_2+\Delta t|<N\\ |m_2-l_1-\Delta t|<N}}^{M}\frac{1}{|m_1m_2l_1l_2|^{1-D}}$$
$$\times\left.\frac{\log|q_2-q_1+\Delta t+m_1-l_2|+\log|q_1-q_2-\Delta t+m_2-l_1|}{|m_1-l_2+m_2-l_1|}\right|_{\partial A}$$

where $\big|_{\partial A}$ means that the antiderivative with respect to q_1 is computed at all values of the boundary of A and always combined via a sum. We observe that the construction of A together with the conditions on q_i, m_i, l_i and j_2 imply that the arguments in the log-function are between 1 and $2N$. Furthermore, for chosen q_2, m_1, l_2 and j_1, there is at most one possible choice for j_2 for which the corresponding summand does not vanish. Thus we have to show that

$$\frac{1}{TN}\sum_{\substack{m_1,m_2,l_1,l_2=1\\ m_1-l_2+m_2-l_1\neq 0}}^{\infty}\sum_{q_2=0}^{N-1}\sum_{\substack{j_2=1\\ 0\leq q_2+m_1-l_2+\Delta t\leq N-1\\ |m_1-l_2+\Delta t|<N\\ |m_2-l_1-\Delta t|<N}}^{M}\frac{1}{|m_1m_2l_1l_2|^{1-D}}\frac{\log(N)}{|m_1-l_2+m_2-l_1|} \tag{6.21}$$

satisfies the bound from (6.17), unformly in $|\Delta t| \geq N$. In fact, due to constraints such as $|m_1 - l_2 + \Delta t| < N$ the expression (6.21) becomes largest for $|\Delta t| = N$, as in this case m_1 and l_2 can jointly be chosen 'small'. Then it follows from the restriction $0 \leq q_2 + m_1 - l_2 + \Delta t \leq N - 1$ on q_2 that there are only $|m_1 - l_2|$ possible choices for q_2 if m_1 and l_2 are chosen. Therefore (6.21) is bounded by

$$\frac{1}{TN} \sum_{\substack{m_1,m_2,l_1,l_2=1 \\ |m_1-l_2+N|<N \\ |m_2-l_1-N|<N \\ m_1-l_2+m_2-l_1\neq 0}}^{\infty} \frac{\log(N)}{|m_1 m_2 l_1 l_2|^{1-D}} \frac{|m_1 - l_2|}{|m_1 - l_2 + m_2 - l_1|}$$

$$= \frac{1}{TN} \sum_{\substack{m_1,m_2,l_1,l_2=1 \\ |m_1-l_2|<2N \\ |m_2-l_1|<2N \\ m_1-l_2+m_2-l_1\neq 0}}^{\infty} \frac{\log(N)}{|m_1 m_2 l_1 l_2|^{1-D}} \frac{|m_1 - l_2|}{|m_1 - l_2 + m_2 - l_1|}$$

which is of order $O(\log(N)^2/(TN^{1-4D}))$ due to Lemma 6.4 c). We have thus shown (6.17) and can restrict ourselves to the case $j_1 = j_2$ in the first term of (6.16), i.e. $V_{T,1}$ equals

$$\begin{aligned}
&\frac{1}{T^2} \sum_{j=1}^{M} \sum_{k_1,k_2=1}^{N/2} \phi_{v_1,\omega_1,T}(u_j, \lambda_{k_1}) \phi_{v_2,\omega_2,T}(u_j, \lambda_{k_2}) \\
&\sum_{\substack{m_1,m_2,l_1,l_2=0 \\ (+)}}^{\infty} \sum_{\substack{q_1,q_2=0 \\ 0\leq q_2+m_1-l_2\leq N-1 \\ 0\leq q_1+m_2-l_1\leq N-1}}^{N-1} \psi_{m_1}(u_j)\psi_{l_1}(u_j)\psi_{m_2}(u_j)\psi_{l_2}(u_j) \\
&\times \frac{1}{(2\pi N)^2} \exp(-i\lambda_{k_1}(q_2 - q_1 + m_1 - l_2)) \exp(i\lambda_{k_2}(q_2 - q_1 + l_1 - m_2)) \\
&+ O\left(\frac{\log(N)^2}{TN^{1-4D}}\right),
\end{aligned}$$

where $(+)$ is a shortcut for $\max(|m_1 - l_2|, |m_2 - l_1|) < N$ and is due to the restrictions on q_i.

Note first that we make an error of order $O(\log(N)^2/(TN^{1-4D}))$ if we skip the conditions on the choice of q_1 and q_2. This follows in a similar way as above, using (6.5), (6.11) and Lemma 6.4 c) once more. Therefore

$$\begin{aligned}
V_{T,1} = &\frac{1}{T^2} \sum_{j=1}^{M} \sum_{k_1,k_2=1}^{N/2} \phi_{v_1,\omega_1,T}(u_j, \lambda_{k_1}) \phi_{v_2,\omega_2,T}(u_j, \lambda_{k_2}) \frac{1}{(2\pi N)^2} \sum_{\substack{m_1,m_2,l_1,l_2=0 \\ (+)}}^{\infty} \sum_{q_1,q_2=0}^{N-1} \\
&\psi_{m_1}(u_j)\psi_{l_1}(u_j)\psi_{m_2}(u_j)\psi_{l_2}(u_j) \\
&\times \exp(-i\lambda_{k_1}(q_2 - q_1 + m_1 - l_2)) \exp(i\lambda_{k_2}(q_2 - q_1 + l_1 - m_2)) \\
&+ O\left(\log(N)^2/(TN^{1-4D})\right).
\end{aligned}$$

By employing the well known identity

$$\frac{1}{N}\sum_{q=0}^{N-1}\exp(-i(\lambda_{k_1}-\lambda_{k_2})q)=\begin{cases}1, & k_1-k_2=lN \text{ with } l\in\mathbb{Z},\\ 0, & \text{else,}\end{cases} \tag{6.22}$$

it can be seen that all terms with $k_1 \neq k_2$ are equal to zero and we therefore get

$$\begin{aligned} V_{T,1} = {} & \frac{1}{T^2}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v_1,\omega_1,T}(u_j,\lambda_k)\phi_{v_2,\omega_2,T}(u_j,\lambda_k)\frac{1}{(2\pi)^2}\sum_{\substack{m_1,m_2,l_1,l_2=0\\(+)}}^{\infty} \\ & \psi_{m_1}(u_j)\psi_{l_1}(u_j)\psi_{m_2}(u_j)\psi_{l_2}(u_j)\exp(-i\lambda_k(m_1-l_2+m_2-l_1)) \\ & + O\left(\log(N)^2/(TN^{1-4D})\right). \end{aligned}$$

The same error arises due to Lemma 6.3, (6.5) and Lemma 6.4 b) if we skip the condition (+). Note that we can proceed completely analogously for the term $V_{T,2}$ with the difference that instead of the right hand side in (6.22) we obtain the corresponding term with $\lambda_{k_1}-\lambda_{k_2}$ replaced by $\lambda_{k_1}+\lambda_{k_2}$. Because of (6.22) we then only have to consider the case $k_1=k_2=N/2$ and therefore the whole term is of order $O(\log(N)^2/(TN^{1-4D}))$. Using the definition of the spectral density the claim follows for V_T^{appr}.

What remains is to show

$$V_T^{true}=V_T^{appr}+O_p\left(\frac{N}{T^2}(N^{4D-1}\log(N)+1)\right),$$

from which the claim follows due to $N=o(T)$. However, the only property of the coefficents $\psi_l(\cdot)$ used in the treatment of $V_{T,1}$ is (6.5). Since (2.9) provides the same property as (6.5) for the original coefficents, we obtain that V_T^{true} equals the final quantity above but with the approximating functions $\psi_l(u_j)$ recplaced by some $\psi_{t_j+c_{j,l}N,T,l}$, $c_{j,l}\in(-1,1)$. Condition (2.5) together with (essentially) Lemma 6.4 c) then yields that we make an error of the order specified above, if we replace $\psi_{t_j+c_{j,l}N,T,l}$ by $\psi_l(t_j+c_{j,l}N/T)$. A Taylor expansion combined with (2.8) then gives the result. □

Proof of c). Assume w.l.o.g. that $(v,\omega):=(v_1,\omega_1)=(v_2,\omega_2)=\cdots=(v_l,\omega_l)$. Using the same replacement of coefficients as in the previous proof we obtain from (2.5), a Taylor expansion and (2.8) that $\mathrm{cum}_l(\hat{E}_T(v,\omega))$ equals

$$\begin{aligned} & \frac{1}{T^l}\sum_{j_1,\ldots,j_l=1}^{M}\sum_{k_1,\ldots,k_l=1}^{N/2}\phi_{v,\omega,T}(u_{j_1},\lambda_{k_1})\cdots\phi_{v,\omega,T}(u_{j_l},\lambda_{k_l})\frac{1}{(2\pi N)^l}\sum_{p_1,\ldots,q_l=0}^{N-1}\sum_{m_1,\ldots,n_l=0}^{\infty} \\ & \mathrm{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1}Z_{t_{j_1}-N/2+1+q_1-n_1},\ldots,Z_{t_{j_l}-N/2+1+p_l-m_l}Z_{t_{j_l}-N/2+1+q_l-n_l}) \\ & \times\psi_{m_1}(u_{j_1})\cdots\psi_{n_l}(u_{j_l})\exp(-i\lambda_{k_1}(p_1-q_1))\cdots\exp(-i\lambda_{k_l}(p_l-q_l))(1+o(1)) \end{aligned}$$

for $l\geq 3$. We define $Y_{i,1}:=Z_{t_{j_i}-N/2+1+p_i-m_i}$ and $Y_{i,2}:=Z_{t_{j_i}-N/2+1+q_i-n_i}$ for $i\in\{1,\ldots,l\}$. Following chapter 2.3 of Brillinger (1981) we obtain

$\text{cum}_l(\hat{E}_T(v,\omega)) = \sum_\nu V_T(\nu)(1+o(1))$, where the sum runs over all indecomposable partitions $\nu = \nu_1 \cup \ldots \cup \nu_l$ with $|\nu_i| = 2$ $(1 \le i \le l)$ of the matrix

$$\begin{matrix} Y_{1,1} & Y_{1,2} \\ \vdots & \vdots \\ Y_{l,1} & Y_{l,2} \end{matrix}$$

and

$$V_T(\nu) := \frac{1}{T^l} \sum_{j_1,\ldots,j_l=1}^{M} \sum_{k_1,\ldots,k_l=1}^{N/2} \phi_{v,\omega,T}(u_{j_1},\lambda_{k_1}) \cdots \phi_{v,\omega,T}(u_{j_l},\lambda_{k_l}) \frac{1}{(2\pi N)^l} \sum_{p_1,\ldots,q_l=0}^{N-1}$$
$$\sum_{m_1,\ldots,n_l=0}^{\infty} \psi_{m_1}(u_{j_1}) \cdots \psi_{n_l}(u_{j_l}) \text{cum}(Y_{i,k};(i,k)\in\nu_1) \cdots \text{cum}(Y_{i,k};(i,k)\in\nu_l)$$
$$\times \exp(-i\lambda_{k_1}(p_1-q_1)) \cdots \exp(-i\lambda_{k_l}(p_l-q_l)).$$

We now fix one indecomposable partition $\tilde{\nu}$ and assume without loss of generality that

$$\tilde{\nu} = \bigcup_{i=1}^{l-1} (Y_{i,1}, Y_{i+1,2}) \cup (Y_{l,1}, Y_{1,2}). \tag{6.23}$$

Because of $\text{cum}(Z_i, Z_j) \neq 0$ for $i \neq j$ we obtain the equations $q_1 = p_l + n_1 - m_l + t_{j_l} - t_{j_1}$ and $q_{i+1} = p_i + n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}}$ for $i \in \{1,\ldots,l-1\}$. Therefore only l variables of the $2l$ variables $p_1, q_1, p_2, \ldots, q_l$ are free to choose and must satisfy the conditions

$$\begin{aligned} &0 \le p_l + n_1 - m_l + t_{j_l} - t_{j_1} \le N-1 \quad \text{and} \\ &0 \le p_i + n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}} \le N-1 \text{ for } i \in \{1,\ldots,l-1\}. \end{aligned} \tag{6.24}$$

Thus we obtain

$$\begin{aligned} V_T(\tilde{\nu}) = \frac{1}{T^l} &\sum_{j_1,\ldots,j_l=1}^{M} \sum_{k_1,\ldots,k_l=1}^{N/2} \phi_{v,\omega,T}(u_{j_1},\lambda_{k_1}) \cdots \phi_{v,\omega,T}(u_{j_l},\lambda_{k_l}) \\ &\times \frac{1}{(2\pi N)^l} \sum_{p_1,\ldots,p_l=0}^{N-1} \sum_{\substack{m_1,\ldots,n_l=0 \\ (6.24)}}^{\infty} \psi_{m_1}(u_{j_1}) \cdots \psi_{n_l}(u_{j_l}) \\ &\times \exp(-i\lambda_{k_1}(p_1 - p_l + m_l - n_1 + t_{j_1} - t_{j_l})) \\ &\prod_{i=2}^{l} \exp(-i\lambda_{k_i}(p_i - p_{i-1} + m_{i-1} - n_i + t_{j_i} - t_{j_{i-1}})). \end{aligned}$$

Note that (6.24) can only be satisfied if $|n_1 - m_l + t_{j_l} - t_{j_1}| < N$ and $|n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}}| < N$ hold for $i \in \{1,2,\ldots,l-1\}$. Using this fact in combination with (6.5) and (6.11) the term above is (up to a constant) bounded by

$$\frac{1}{T^l}\sum_{j_1=1}^{M}\sum_{\substack{m_1,n_1,\ldots,m_l,n_l=0\\ m_i,n_i\neq 0}}^{\infty}\sum_{\substack{j_2,\ldots,j_l=1\\ |n_{i+1}-m_i+t_{j_i}-t_{j_{i+1}}|<N\\ |n_1-m_l+t_{j_l}-t_{j_1}|<N}}^{M}\frac{1}{|m_1|^{1-d}}\cdots\frac{1}{|n_l|^{1-d}}$$

$$\sum_{\substack{p_1,p_2,\ldots,p_l=0\\ |p_i-p_{i-1}+m_{i-1}-n_i+t_{j_i}-t_{j_{i-1}}|<N/2}}^{N-1}\frac{1}{|p_1-p_l+m_l-n_1+t_{j_1}-t_{j_l}|}$$

$$\prod_{i=2}^{l}\frac{1}{|p_i-p_{i-1}+m_{i-1}-n_i+t_{j_i}-t_{j_{i-1}}|}\prod_{i=1}^{l}1(p_i\notin\{z_{i1},z_{i2}\})$$

where z_{i1}, z_{i2} are the p_i for which the denominator vanishes, i.e. $z_{i1}=p_{i-1}+n_i-m_{i-1}+t_{j_{i-1}}-t_{j_i}$ and $z_{i2}=p_{i+1}+m_i-n_{i+1}+t_{j_{i+1}}-t_{j_i}$ for $i=\{1,\ldots,l\}$, where we identified 0 with l and $l+1$ with 1. Note that the cases with $p_i=z_{ij}$ for a $j\in\{1,2\}$ or $|p_i-p_{i-1}+m_{i-1}-n_i+t_{j_i}-t_{j_{i-1}}|\geq N/2$ are again of smaller or equal order. Recall the treatment of (6.20). If we set $A_i=[0,N-1]\backslash([z_{i1}-1,z_{i1}+1]\cup[z_{i2}-1,z_{i2}+1])$ for $i=\{1,\ldots,l\}$, the final line of the previous display can be bounded by

$$\sum_{p_l=0}^{N-1}\int_{A_1\times\ldots\times A_{l-1}}\frac{1}{|p_1-p_l+m_l-n_1+t_{j_1}-t_{j_l}|}$$

$$\prod_{i=2}^{l}\frac{1}{|p_i-p_{i-1}+m_{i-1}-n_i+t_{j_i}-t_{j_{i-1}}|}d(p_1,\ldots,p_{l-1})$$

$$\leq\sum_{p_l=0}^{N-1}\int_{A_2\times\ldots\times A_{l-1}}$$

$$\left.\frac{\log|p_1-p_l+m_l-n_1+t_{j_1}-t_{j_l}|+\log|p_2-p_1+m_1-n_2+t_{j_2}-t_{j_1}|}{|p_2-p_l+t_{j_2}-t_{j_l}+m_l-n_1+m_1-n_2|}\right|_{\partial A_1}$$

$$\times\prod_{i=3}^{l}\frac{1}{|p_i-p_{i-1}+m_{i-1}-n_i+t_{j_i}-t_{j_{i-1}}|}d(p_2,\ldots,p_{l-1}).$$

where we considered partial fractions again and with the same notation as before. The conditions on p_i,m_i,n_i and j_i imply that the arguments of the log-functions are between 1 and $2N$, so also smaller than $2lN$. Thus the above term bounded by

$$\sum_{p_l=0}^{N-1}\int_{A_2\times\ldots\times A_{l-1}}\frac{\log(2lN)}{|p_2-p_l+t_{j_2}-t_{j_l}+m_l-n_1+m_1-n_2|}$$

$$\prod_{i=3}^{l}\frac{1}{|p_i-p_{i-1}+m_{i-1}-n_i+t_{j_i}-t_{j_{i-1}}|}d(p_2,\ldots,p_{l-1}).$$

Using this argumentation also in the integration over $p_2, \dots, p_{l-1}$, we can bound $V_T(\tilde{\nu})$ (up to a constant) by

$$\frac{1}{T^l} \sum_{j_1=1}^{M} \sum_{\substack{m_1,n_1,\dots,m_l,n_l=1 \\ m_1-n_1+\cdots+m_l-n_l\neq 0}}^{\infty} \sum_{\substack{j_2,\dots,j_l=1 \\ |n_{i+1}-m_i+t_{j_i}-t_{j_{i+1}}|<N \\ |n_1-m_l+t_{j_l}-t_{j_1}|<N}}^{M} \frac{1}{|m_1|^{1-d}} \cdots \frac{1}{|n_l|^{1-d}} \frac{1}{|m_1-n_1+\cdots+m_l-n_l|} \sum_{p_1=0}^{N-1} \log(2lN)^{l-1}$$

where all the differences of p_i- and t_{j_i}-terms vanish in a telescoping sum. Note that for T large enough, the conditions

$$|n_{i+1} - m_i + t_{j_i} - t_{j_{i+1}}| < N \quad \text{and} \quad |n_1 - m_l + t_{j_l} - t_{j_1}| < N \tag{6.25}$$

can only be satisfied if $|n_{i+1} - m_i| \leq 2T$ for $i \in \{1, \dots, l\}$ where we identified $l+1$ with 1. Therefore the above term is smaller or equal to

$$\frac{1}{T^l} \sum_{j_1=1}^{M} \sum_{\substack{m_1,n_1,\dots,m_l,n_l=1 \\ (+)_{\neq,l}}}^{\infty} \sum_{\substack{j_2,\dots,j_l=1 \\ (6.25)}}^{M} \frac{1}{|m_1|^{1-d}} \cdots \frac{1}{|n_l|^{1-d}} \frac{1}{|m_1-n_1+\cdots+m_l-n_l|} \sum_{p_1=0}^{N-1} \log(2lN)^{l-1}$$

where $(+)_{\neq,l}$ was defined in Lemma 6.4 d) and we now have $n = 2T$. As in the proof of part b), it can be seen that if j_1, m_i, n_i are chosen, there are only finitely many possible choices for $j_2, \dots, j_l$ because of the conditions (6.25). By using this and Lemma 6.4 d), we finally obtain

$$\begin{aligned} V_T(\tilde{\nu}) &= O(T^{1-l} \log(N)^{l-1} \log(T) T^{2Dl-4D}) \\ &= O\left(T^{(1-4D)-l(1/2-2D)-l/2} \log(T)^l\right) \end{aligned}$$

which is of order $o(1/T^{l/2})$ for $l \geq 3$ and $D < 1/4$. □

Proof of d). Recall that $\hat{G}_T(v,\omega) - C_T(v,\omega) = \hat{E}_T(v,\omega) - C_T(v,\omega) - E_T(v,\omega)$. Analogously to the proof of Theorem 5.1 in Preuß et al. (2012) the claim can be proven by finding appropriate bounds for cumulants instead of moments. We begin with the first cumulant, as only in this case the non-random terms play a role. Regarding the other cumulants, we show

$$|\mathrm{cum}_l(\hat{E}_T(v_1,\omega_1) - \hat{E}_T(v_2,\omega_2))| \leq (2l)! C^l \rho_{2,T,D}\left((v_1,\omega_1),(v_2,\omega_2)\right)^l T^{-l/2} \tag{6.26}$$

later only for even l, as the general result can be obtained similarly.

Let us start with the first cumulant. Recall that

$$E[|X|] \leq \sqrt{\mathrm{Var}(X) + \mathbb{E}[X]^2} \leq \sqrt{\mathrm{Var}(X)} + |\mathbb{E}[X]|$$

for any random variable X. Therefore, in order to bound

$$\mathbb{E}|\hat{G}_T(v_1, \omega_1) - C_T(v_1, \omega_1) - (\hat{G}_T(v_2, \omega_2) - C_T(v_2, \omega_2))|$$

we have to discuss two terms, and since the variance equals the second cumulant, the bound for the first term will follow from (6.26) for $l = 2$. Furthermore, it follows from the proof of Theorem 6.1 a) that

$$\mathbb{E}[\hat{G}_T(v, \omega) - C_T(v, \omega)]$$

consists of two error terms which are due to the approximation (2.5). We discuss the one corresponding to the third summand in (6.9) only, which is

$$\frac{1}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u,\lambda)\frac{1}{2\pi N}\sum_{\substack{l,m=0\\|l-m|<N}}^{\infty}\sum_{\substack{q=0\\0\leq q+m-l\leq N-1}}^{N-1}\psi_m\left(\frac{e_{j,N}+q}{T}\right)$$
$$\left(\psi_{e_{j,N}+q+m-l,T,l} - \psi_l\left(\frac{e_{j,N}+q+m-l}{T}\right)\right)\exp(-i\lambda_k(m-l)). \qquad (6.27)$$

We first consider the sum over m. Note that, if l and q are fixed, m has to lie between $\max(l-q, 0)$ and $l+N-1-q$. For $l < q+1$, we obtain with (2.6) and Theorem 2.6 in Chapter V of Zygmund (1959) that

$$\Big|\sum_{m=0}^{l+N-1-q}\psi_m\left(\frac{e_{j,N}+q}{T}\right)\exp(-i\lambda_k m)\Big| \leq C\lambda_k^{-D} \qquad (6.28)$$

holds. For $l \geq q+1$ we get with equation (2.9) from Chapter V in Zygmund (1959) that

$$\begin{aligned}\Big|\sum_{m=l-q}^{l+N-1-q}\psi_m\left(\frac{e_{j,N}+q}{T}\right)\exp(-i\lambda_k m)\Big| &\leq C(l-q)^{D-1}\lambda_k^{-1}\\ &\leq C(l-q)^{D-1}N^{1-2D}\lambda_k^{-2D}.\end{aligned}$$

If we employ (2.5) thereafter and use approximating integrals in the summation over l we get that (6.27) is smaller than a positive constant times

$$\frac{1}{T^2}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u,\lambda)\frac{1}{2\pi N}\sum_{q=0}^{N-1}\log(q)\Big(\lambda_k^{-D}q^D + N^{1-2D}\lambda_k^{-2D}q^{2D-1}\Big).$$

By bounding the sum over q again through approximating integrals, we finally obtain that (6.27) is smaller than

$$\frac{CN^D\log(N)}{T^2}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u,\lambda)\frac{1}{\lambda^{2D}}$$

for some suitable constant $C \in I\!R$, yielding the claim for $l = 1$ because of $N = o(T)$ and $D < 1/2$.

In order to simplify technical arguments for $l \geq 2$ we furthermore define $\phi_{v,\omega,T}(u,\lambda) := \phi_{v,\omega,T}(u,-\lambda)$ for $u \in [0,1]$ and $\lambda \in [-\pi, 0]$. Due to the symmetry of $I_N(u,\lambda)$ in λ we then obtain that the l-th cumulant of $\hat{E}_T(v_1,\omega_1) - \hat{E}_T(v_2,\omega_2))$ is given by

$$\frac{1}{2^l T^l} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,\dots,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} (\phi_{v_1,\omega_1,T}(u_{j_1},\lambda_{k_1}) - \phi_{v_2,\omega_2,T}(u_{j_1},\lambda_{k_1})) \cdots$$
$$\cdots (\phi_{v_1,\omega_1,T}(u_{j_l},\lambda_{k_l}) - \phi_{v_2,\omega_2,T}(u_{j_l},\lambda_{k_l}))$$
$$\times \frac{1}{(2\pi N)^l} \sum_{p_1,\dots,q_l=0}^{N-1} \sum_{m_1,\dots,n_l=0}^{\infty} \psi_{m_1}(u_{j_1}) \cdots \psi_{n_l}(u_{j_l}) \exp(-i\lambda_{k_1}(p_1 - q_1)) \cdots$$
$$\cdots \exp(-i\lambda_{k_l}(p_l - q_l))$$
$$\times \operatorname{cum}(Z_{t_{j_1}-N/2+1+p_1-m_1} Z_{t_{j_1}-N/2+1+q_1-n_1}, \dots$$
$$\dots, Z_{t_{j_l}-N/2+1+p_l-m_l} Z_{t_{j_l}-N/2+1+q_l-n_l})(1 + o(1))$$

Set $\phi_{1,2,T}(u,\lambda) := \phi_{v_1,\omega_1,T}(u,\lambda) - \phi_{v_2,\omega_2,T}(u,\lambda)$. We restrict ourselves again to the indecomposable partition $\tilde{\nu}$ defined in (6.23) and call the corresponding summand $V_{2,T}(\tilde{\nu})$. Then as in the proof of Theorem 5.1 in Preuß et al. (2012) we see that

$$0 \leq p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \leq N - 1 \quad \text{for } i \in \{1, 3, 5, \dots, l-3, l-1\} \tag{6.29}$$

must be satisfied and that $V_{2,T}(\tilde{\nu})$ is bounded by $\sqrt{J_{1,T} J_{2,T}}$ with

$$J_{1,T} = \frac{1}{2^l T^l} \sum_{j_1,\dots,j_l=1}^{M} \sum_{k_1,k_3,\dots,k_{l-1}=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi^2_{1,2,T}(u_{j_1},\lambda_{k_1}) \phi^2_{1,2,T}(u_{j_3},\lambda_{k_3})^2 \cdots$$
$$\cdots \phi^2_{1,2,T}(u_{j_{l-1}},\lambda_{k_{l-1}})^2 \frac{1}{(2\pi N)^l} \sum_{p_1,p_3,\dots,p_{l-1}=0}^{N-1} \sum_{\tilde{p}_1,\tilde{p}_3,\dots,\tilde{p}_{l-1}=0}^{N-1}$$
$$\sum_{\substack{m_1,n_1,m_3,n_3,\dots,m_{l-1},n_{l-1}=0 \\ (6.29)}}^{\infty} \sum_{\substack{\tilde{m}_1,\tilde{n}_1,\tilde{m}_3,\tilde{n}_3,\dots,\tilde{m}_{l-1},\tilde{n}_{l-1}=0 \\ \widetilde{(6.29)}}}^{\infty}$$
$$\exp(-i\lambda_{k_1}(p_1 - \tilde{p}_1)) \exp(-i\lambda_{k_3}(p_3 - \tilde{p}_3)) \cdots \exp(-i\lambda_{k_{l-1}}(p_{l-1} - \tilde{p}_{l-1}))$$
$$\psi_{m_1}(u_{j_2})\psi_{n_1}(u_{j_1}) \cdots \psi_{m_{l-1}}(u_{j_l})\psi_{n_{l-1}}(u_{j_{l-1}})$$
$$\psi_{\tilde{m}_1}(u_{j_2})\psi_{\tilde{n}_1}(u_{j_1}) \cdots \psi_{\tilde{m}_{l-1}}(u_{j_l})\psi_{\tilde{n}_{l-1}}(u_{j_{l-1}})$$
$$\sum_{k_2,k_4,\dots,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \exp(-i\lambda_{k_2}(\tilde{p}_1 - p_1 + n_1 - m_1 + \tilde{m}_1 - \tilde{n}_1))$$
$$\exp(-i\lambda_{k_4}(\tilde{p}_3 - p_3 + n_3 - m_3 + \tilde{m}_3 - \tilde{n}_3)) \cdots$$
$$\cdots \exp(-i\lambda_{k_l}(\tilde{p}_{l-1} - p_{l-1} + n_{l-1} - m_{l-1} + \tilde{m}_{l-1} - \tilde{n}_{l-1})) \tag{6.30}$$

and $J_{2,T}$ being defined for even p_i, m_i, n_i. Here, the condition $\widetilde{(6.29)}$ says that (6.29) holds but with the p_i, m_i, n_i replaced by $\tilde{p}_i, \tilde{m}_i, \tilde{n}_i$. The identity (6.22) implies that in (6.30) the restrictions

$$\begin{aligned}&\tilde{p}_i = p_i + m_i - n_i + \tilde{n}_i - \tilde{m}_i \quad \text{and}\\ &0 \le p_i + m_i - n_i + \tilde{n}_i - \tilde{m}_i \le N-1 \text{ for odd } i\end{aligned} \tag{6.31}$$

must be fulfilled and that $J_{1,T}$ therefore equals

$$\frac{1}{(4\pi)^l T^l N^{l/2}} \sum_{j_1,\ldots,j_l=1}^{M} \sum_{k_1,k_3,\ldots,k_{l-1}=-\lfloor (N-1)/2\rfloor}^{N/2} \phi_{1,2,T}^2(u_{j_1},\lambda_{k_1})\phi_{1,2,T}^2(u_{j_3},\lambda_{k_3})\cdots$$
$$\phi_{1,2,T}^2(u_{j_{l-1}},\lambda_{k_{l-1}}) \sum_{p_1,\ldots,p_{l-1}=0}^{N-1} \sum_{\substack{m_1,\ldots,n_{l-1}=0\\ (6.29)}}^{\infty} \sum_{\substack{\tilde{m}_1,\ldots,\tilde{n}_{l-1}=0\\ (6.31)}}^{\infty} \psi_{m_1}(u_{j_2})\cdots\psi_{n_{l-1}}(u_{j_{l-1}})$$
$$\psi_{\tilde{m}_1}(u_{j_2})\cdots\psi_{\tilde{n}_{l-1}}(u_{j_{l-1}})\exp(-i\lambda_{k_1}(n_1 - m_1 + \tilde{m}_1 - \tilde{n}_1))\cdots$$
$$\cdots\exp(-i\lambda_{k_{l-1}}(n_{l-1} - m_{l-1} + \tilde{m}_{l-1} - \tilde{n}_{l-1})).$$

A factorisation yields $J_{1,T} = L_{1,T} \times L_{3,T} \times \cdots \times L_{l-1,T}$ with

$$\begin{aligned}L_{i,T} :=& \frac{1}{16\pi^2 T^2} \sum_{j_i=1}^{M} \sum_{k_i=-\lfloor (N-1)/2\rfloor}^{N/2} \phi_{1,2,T}^2(u_{j_i},\lambda_{k_i}) \sum_{\substack{m_i,n_i,\tilde{m}_i,\tilde{n}_i=0\\ |m_i-n_i+\tilde{n}_i-\tilde{m}_i|<N}}^{\infty}\\ &\frac{1}{N} \sum_{\substack{p_i=0\\ 0\le p_i+m_i-n_i+\tilde{n}_i-\tilde{m}_i\le N-1}}^{N-1} \sum_{\substack{j_{i+1}=1\\ 0\le p_i+m_i-n_i+t_{j_i}-t_{j_{i+1}}\le N-1}}^{M} \psi_{m_i}(u_{j_{i+1}})\psi_{n_i}(u_{j_i})\\ &\psi_{\tilde{m}_i}(u_{j_{i+1}})\psi_{\tilde{n}_i}(u_{j_i})\exp(-i\lambda_{k_i}(n_i - m_i + \tilde{m}_i - \tilde{n}_i))\\ =& \frac{1}{16\pi^2 T^2} \sum_{j_i=1}^{M} \sum_{k_i=-\lfloor (N-1)/2\rfloor}^{N/2} \phi_{1,2,T}^2(u_{j_i},\lambda_{k_i}) h_{sup,j_i,N}(\lambda_k)\end{aligned}$$

with $h_{sup,j_i,N}(\lambda)$ being defined implicitly. In the following we will show that there exist a constant $C \in \mathbb{R}$ such that

$$-C^2/\lambda_k^{4D} \le h_{sup,j_i,N}(\lambda_k) \le C^2/\lambda_k^{4D} \tag{6.32}$$

for all j_i, N, λ_k, which then yields that each $|L_{i,T}|$ is up to a constant bounded by

$$\frac{1}{T^2}\sum_{j=1}^{M}\sum_{k=1}^{N/2} \phi_{1,2,T}(u_j,\lambda_k)^2 \frac{1}{\lambda_k^{4D}}.$$

This implies $J_{1,T} \le C^l \rho_{2,T,D}((v_1,\omega_1),(v_2,\omega_2))^l T^{-l/2}$ and since the same upper bound is obtained for $J_{2,T}$ the claim then follows analogously to the proof of

Theorem 5.1 in Preuß et al. (2012) by employing that $(2l)!2^l$ is an upper bound for the number of indecomposable partitions.

Concerning the proof of (6.32) we restrict ourselves to the right inequality and assume without loss of generality that $d(u_{j_i}) = D$. It turns out that this case is the 'worst' one in the sense that it yields the largest possible upper bound. By setting $l = n_i - m_i + \tilde{m}_i - \tilde{n}_i$ and employing (2.6) we obtain

$$h_{sup,j_i,N}(\lambda) = \left(\sum_{l=-N+1}^{N-1} b_{j_i,N,l} \exp(-i\lambda l) \right) (1 + o(1)),$$

where the $o(1)$ term is uniformly bounded in j_i, N, λ and where

$$b_{j_i,N,l} = \sum_{n_i=0}^{\infty} \frac{1}{N} \sum_{\substack{p_i=0 \\ 0 \le p_i - l \le N-1}}^{N-1} \sum_{\substack{m_i=0 \\ b_{1,N}(j_i,p_i) \le m_i - n_i \le b_{2,N}(j_i,p_i)}}^{\infty} \sum_{\tilde{m}_i = l + m_i - n_i}^{\infty} \frac{a(u_{j_i})^2 a(u_{g(j_i,p_i,m_i-n_i)})^2}{I(n_i(\tilde{m}_i - m_i + n_i - l))^{1-d(u_{j_i})} I(m_i \tilde{m}_i)^{1-d(u_{g(j_i,p_i,m_i-n_i)})}}.$$

Here, $b_{1,N}(j_i, p_i)$ and $b_{2,N}(j_i, p_i)$ are chosen to ensure that for given j_i, p_i there exists a $j_{i+1} \in \{1, \dots, M\}$ such that $0 \le p_i + m_i - n_i + t_{j_i} - t_{j_{i+1}} \le N - 1$ can be satisfied. For an admissible pair (m_i, n_i), $g(j_i, p_i, m_i - n_i)$ then denotes the corresponding j_{i+1}. Note that there exists at most one such j_{i+1} for a given (m_i, n_i), and if p_i is fixed and $|m_i - n_i| < T/2$ then there always exist such an j_{i+1}, implying that $b_{1,N}(j_i, p_i) \le -T/2$ and $b_{2,N}(j_i, p_i) \ge T/2$ for all j_i, p_i. If we then make use of the fact that $a(\cdot)$ is strictly positive and replace the sums in $b_{j_i,N,a}$ through approximating integrals we obtain

$$b_{j_i,N,l} = \frac{N - |l|}{N} \frac{\tilde{b}_{j_i,N,l}}{I(l)^{1-4D}}$$

with $0 < C_1 \le \tilde{b}^1_{j_i,N,l} \le C_2 < \infty$ for constants C_1, C_2 uniformly in j_i, N, l. The claim now follows as in the treatment of (6.27) by employing the periodicity of $\exp(x)$ and approximating integrals. See the second section of Chapter V in Zygmund (1959) for a detailed analysis of such terms. □

6.2. *Proof of Theorem 6.2*

The proof works in the same way as the proof of Theorem 6.1 but by employing Lemma 4.3 instead of (2.6) in order to keep error terms uniformly small in probability.

Proof of a). At first note that the coeffcients in the $MA(\infty)$ representations (4.8) and (4.11) do not depend on the time. Thus, if we write $I_N^*(u, \lambda)$ for the bootstrap analogon of $I_N(u, \lambda)$, we obtain

$$\mathbb{E}\left(\hat{E}_T^*(v, \omega) 1_{A_T(\alpha)} \middle| X_{1,T}, \dots, X_{T,T} \right) \tag{6.33}$$

$$= \mathbb{E}\big(\frac{1_{A_T(\alpha)}}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)I_N^*(u_j,\lambda_k)|X_{1,T},\dots,X_{T,T}\big)$$

$$= \frac{1_{A_T(\alpha)}}{T}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v,\omega,T}(u_j,\lambda_k)\frac{1}{2\pi N}\sum_{p,q=0}^{N-1}\sum_{l,m=0}^{\infty}\hat{\psi}_{l,p}\hat{\psi}_{m,q} \quad (6.34)$$
$$\mathbb{E}(Z^*_{t_j-N/2+1+p-m}Z^*_{t_j-N/2+1+q-l})\exp(-i\lambda_k(p-q)).$$

The $\hat{\psi}_{l,p}$ possess no time dependence, thus the above expression equals zero by definition of $\phi_{v,\omega,T}$. The same result holds for $\hat{G}^*_{T,a}(v,\omega)$. □

Proof of b). Because of part a) we obtain

$$\mathrm{Var}\Big((\hat{E}^*_T(v,\omega)-\hat{E}^*_{T,a}(v,\omega))1_{A_T(\alpha)}\Big)$$
$$= \mathbb{E}\Big(\mathrm{Var}(\hat{E}^*_T(v,\omega)-\hat{E}^*_{T,a}(v,\omega)|X_{1,T},\dots,X_{T,T})1_{A_T(\alpha)}\Big)$$
$$= \frac{1}{T^2}\sum_{j_1,j_2=1}^{M}\sum_{k_1,k_2=1}^{N/2}\phi_{v_1,\omega_1,T}(u_{j_1},\lambda_{k_1})\phi_{v_2,\omega_2,T}(u_{j_2},\lambda_{k_2})\frac{1}{(2\pi N)^2}$$
$$\sum_{p_1,p_2,q_1,q_2=0}^{N-1}\sum_{m_1,m_2,l_1,l_2=0}^{\infty}\mathbb{E}(\hat{\psi}_{m_1,l_1,m_2,l_2,p}1_{A_T(\alpha)})$$
$$\times\ \mathrm{cum}(Z^*_{e_{j_1,N}+p_1-m_1}Z^*_{e_{j_1,N}+q_1-l_1},Z^*_{e_{j_2,N}+p_2-m_2}Z^*_{e_{j_2,N}+q_2-l_2})$$
$$\times\exp(-i\lambda_{k_1}(p_1-q_1))\exp(-i\lambda_{k_2}(p_2-q_2))$$

with $\hat{\psi}_{m_1,l_1,m_2,l_2,p}=(\hat{\psi}_{m_1,p}\hat{\psi}_{l_1,p}-\tilde{\psi}_{m_1}\tilde{\psi}_{l_1})(\hat{\psi}_{m_2,p}\hat{\psi}_{l_2,p}-\tilde{\psi}_{m_2}\tilde{\psi}_{l_2})$. By using

$$\hat{\psi}_{m_1,p}\hat{\psi}_{l_1,p}-\tilde{\psi}_{m_1}\tilde{\psi}_{l_1}=(\hat{\psi}_{m_1,p}-\tilde{\psi}_{m_1})\tilde{\psi}_{l_1}+(\hat{\psi}_{l_1,p}-\tilde{\psi}_{l_1})\hat{\psi}_{m_1,p}$$

and the analogue for $\hat{\psi}_{m_2,p}\hat{\psi}_{l_2,p}-\tilde{\psi}_{m_2,p}\tilde{\psi}_{l_2,p}$, we can divide the above expression into the sum of four terms. For the sake of brevity details are presented only for the first one. By using (6.15) the corresponding summand splits into two terms and we restrict ourselves to the first one which we denote with $V^*_{T,1}$. As in the proof of Theorem 6.1 b) we then obtain an error term $V^{j_1\neq j_2,*}_{T,1}$ which is defined as $V^{j_1\neq j_2}_{T,1}$ but with the $\psi_{m_1}(u_{j_1})\psi_{l_1}(u_{j_1})\psi_{m_2}(u_{j_2})\psi_{l_2}(u_{j_2})$ replaced by $\mathbb{E}\Big((\hat{\psi}_{m_1,p}-\tilde{\psi}_{m_1})(\hat{\psi}_{m_2,p}-\tilde{\psi}_{m_2})\tilde{\psi}_{l_1}\tilde{\psi}_{l_2}1_{A_T(\alpha)}\Big)$. In the following we will demonstrate that

$$V^{j_1\neq j_2,*}_{T,1}=O\left(p^{10}\log(T)^2\log(N)^2N^{\max(4D-1,0)+\alpha}T^{-2}\right).$$

The proof is similar to the one of (6.17) up to employing (6.3). Let us demonstrate this concept in the treatment of $V^{j_1\neq j_2,*}_{T,1,3}$ which is bounded by

$$\frac{1}{(2\pi T)^2}\sum_{j_1=1}^{M}\sum_{m_1,m_2,l_1,l_2=0}^{\infty}\sum_{q_2=0}^{N-1}\sum_{\substack{j_2=1\\ j_2\neq j_1\\ 0\le q_2+m_1-l_2+\Delta t\le N-1}}^{M}$$

$$\mathbb{E}|(\hat{\psi}_{m_1,p} - \tilde{\psi}_{m_1})(\hat{\psi}_{m_2,p} - \tilde{\psi}_{m_2})\tilde{\psi}_{l_1}\tilde{\psi}_{l_2}1_{A_T(\alpha)}| \sum_{\substack{q_1=0 \\ 0\le q_1+m_2-l_1-\Delta t\le N-1 \\ (6.18)}}^{N-1}$$

$$\Big|\frac{1}{N}\sum_{k_1=1}^{N/2}\phi_{v_1,\omega_1,T}(u_{j_1},\lambda_{k_1})\exp(-i\lambda_{k_1}(q_2-q_1+\Delta t+m_1-l_2))\Big|$$

$$\Big|\frac{1}{N}\sum_{k_2=1}^{N/2}\phi_{v_2,\omega_2,T}(u_{j_2},\lambda_{k_2})\exp(-i\lambda_{k_2}(q_1-q_2-\Delta t+m_2-l_1))\Big|;$$

compare with (6.19). In the proof of Theorem 6.1 b) we have shown that $V_{T,1,3}^{j_1\neq j_2}$ is of order $O(\log(N)^2/(TN^{1-4D}))$ by employing (6.5). Here we use (6.3) instead and combine it with the fact that $|\hat{D}-D|<\alpha/4$ on $A_T(\alpha)$ to obtain

$$|\hat{\psi}_{l,p} - \tilde{\psi}_l| \le Cp^5\log(T)T^{-1/2}|l|^{\alpha/4+D-1} \quad \forall l\in\mathbb{N}. \tag{6.35}$$

This together with (6.5) and Assumption 4.2 implies

$$|\hat{\psi}_{l,p}| \le C|l|^{\alpha/4+D-1} \quad \forall l,p\in\mathbb{N}. \tag{6.36}$$

Thus the role of D is played by $D+\alpha/4$ now, and using (6.35) and (6.36) instead of (6.5) we obtain

$$\begin{aligned} V_{T,1,3}^{j_1\neq j_2,*} &\le Cp^{10}\log(T)^2T^{-1}\times\log(N)^2T^{-1}N^{4D+\alpha-1} \\ &\le Cp^{10}\log(T)^2\log(N)^2N^{\max(4D-1,0)+\alpha}T^{-2}. \end{aligned}$$

Similarly, the subsequent steps in the proof of Theorem 6.1 b) reveal that $V_{T,1}^*$ becomes

$$\frac{1}{T^2}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi_{v_1,\omega_1,T}(u_j,\lambda_k)\phi_{v_2,\omega_2,T}(u_j,\lambda_k)\frac{1}{(2\pi)^2}\sum_{\substack{m_1,m_2,l_1,l_2=0 \\ (+)}}^{\infty}$$

$$\exp(-i\lambda_k(m_1-l_2+m_2-l_1))\mathbb{E}\Big((\hat{\psi}_{m_1,p}-\tilde{\psi}_{m_1})(\hat{\psi}_{m_2,p}-\tilde{\psi}_{m_2})\tilde{\psi}_{l_1}\tilde{\psi}_{l_2}1_{A_T(\alpha)}\Big)$$

$$+\,O\Big(p^{10}\log(T)^2\log(N)^2N^{\max(4D-1,0)+\alpha}T^{-2}\Big).$$

In the proof of Theorem 6.1 b), the analogue of the first quantity on the right hand side above is the main term contributing to the variance. Here, however, it is of the same order as the error terms. This can be seen using (6.35) and (6.36) again plus Lemma 6.4 c). □

Proof of c). Since the expectation is zero by construction, it is clearly sufficient to focus on the higher cumulants only. If we employ (6.35) and (6.36) as in the proof of part b) and follow the arguments in the proof of Theorem 6.1 d), we obtain

$$\mathbb{E}\Big(|(\hat{E}_T^*(v_1,\omega_1)-\hat{E}_{T,a}^*(v_1,\omega_1))-(\hat{E}_T^*(v_2,\omega_2)-\hat{E}_{T,a}^*(v_2,\omega_2))|^k1_{A_T(\alpha)}\Big)$$

$$\leq (2k)!C^k \rho_{2,T,D+\alpha/4}\left((v_1,\omega_1),(v_2,\omega_2)\right)^k \left(p^{10}\log(T)^2 T^{-2}\right)^{k/2},$$

where $\rho_{2,T,D+\alpha/4}(\cdot,\cdot)$ corresponds to the metric defined in (6.1) but with D replaced by $D+\alpha/4$ due to $|\hat{D}-D|\leq \alpha/4$. The claim then follows from

$$\rho_{2,T,D+\alpha/4}\left((v_1,\omega_1),(v_2,\omega_2)\right) \leq C\tilde{\rho}\left((v_1,\omega_1),(v_2,\omega_2)\right)\sqrt{N^{\max(4D-1,0)+\alpha}}. \quad \square$$

6.3. An auxiliary result used for the bootstrap

Lemma 6.6. *Suppose that there are random variables which satisfy the conditions*

- $\hat{W}^*_{T,a} \stackrel{\mathcal{D}}{=} \hat{W}_{T,a}$,
- $\hat{W}^*_T = \hat{W}^*_{T,a} + Z^*_T$ *with* $Z^*_T = o_P(1)$,
- $\hat{W}_T = \hat{W}_{T,a} + Z_T$ *with* $Z_T = o_P(1)$.

Let w_T *denote the* γ *quantile of* $\hat{W}^*_T$*. Then for any* $\varepsilon,\delta>0$ *there exists some* $T_0>0$ *such that*

$$P(\hat{W}_T \leq w_T - \delta) - \varepsilon < \gamma < P(\hat{W}_T \leq w_T + \delta) + \varepsilon$$

for any $T \geq T_0$.

Proof. We have

$$\begin{aligned}\gamma \leq P(\hat{W}^*_T \leq w_T + \delta/3) &= P(\hat{W}^*_{T,a} + Z^*_T \leq w_T + \delta/3)\\ &\leq P(\hat{W}^*_{T,a} \leq w_T + 2\delta/3) + P(|\hat{Z}^*_T| \geq \delta/3).\end{aligned}$$

Furthermore,

$$\begin{aligned}P(\hat{W}^*_{T,a} \leq w_T + 2\delta/3) = P(\hat{W}_{T,a} \leq w_T + 2\delta/3) &= P(\hat{W}_T - Z_T \leq w_T + 2\delta/3)\\ &\leq P(\hat{W}_T \leq w_T + \delta) + P(|\hat{Z}_T| \geq \delta/3).\end{aligned}$$

Finally, choose T_0 large enough to secure that both $P(|\hat{Z}^*_T| \geq \delta/3) < \varepsilon/2$ and $P(|\hat{Z}_T| \geq \delta/3) < \varepsilon/2$ hold for any $T \geq T_0$. This gives the upper bound. The lower one follows in the same way. $\square$

6.4. Proof of the results from the main corpus

Proof of Theorem 2.2. Without loss of generality we prove the claim for the case $q=0$ only. The extension to $q>0$ is then straightforward. Following the arguments from the proof of Proposition 2.3 in Dahlhaus and Polonik (2009) we obtain that the process $X_{t,T}$ possesses an MA(∞) representation with innovations $(1-B)^{-d(t/T)}Z_t$ and coefficents

$$a_{0,t,T}=1, \qquad a_{l,t,T} = \Big(\prod_{j=0}^{l-1} \mathbf{a}\Big(\frac{t-j}{T}\Big)\Big)_{11}, \quad l \geq 1,$$

where

$$\mathbf{a}(u) := \begin{pmatrix} -a_1(u) & -a_2(u) & \cdots & \cdots & -a_p(u) \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

By setting $\eta_0(u) \equiv 1$ and

$$\eta_l(u) = \frac{\Gamma(l + d(u))}{\Gamma(d(u))\Gamma(l+1)}, \quad l \geq 1, \tag{6.37}$$

we can invert the operator $(1 - B)^{d(t/T)}$ in order to obtain an MA(∞) representation of the process $X_{t,T}$ with innovations Z_t and coefficents

$$\psi_{l,t,T} = \sum_{k=0}^{l} a_{k,t,T} \eta_{l-k}(t/T).$$

Finally, we set $\tilde{a}_l(u) = (\mathbf{a}(u)^l)_{11}$ and define

$$\psi_l(u) = \sum_{k=0}^{l} \tilde{a}_k(u) \eta_{l-k}(u).$$

In the stationary case, such processes have been investigated in Kokoszka and Taqqu (1995), where it has been shown that the equivalent of (2.6) holds true for each fixed u; cf. their Corollary 3.1. Uniformity of this approximation follows from our assumptions on the coefficients easily. Furthermore, it is straightforward to see that (2.8) and (2.9) are fulfilled, so Assumption 2.1 it satisfied up to (2.5). If we employ the fact that there exists a constant $C \in \mathbb{R}$ such that $\sup_{u \in [0,1]} |\eta_l(u)| \leq C I(l)^{1-D}$, we obtain

$$\begin{aligned} &|\psi_{l,t,T} - \psi_l(t/T)| \\ &\leq C \sum_{k=0}^{l-1} \frac{1}{(l-k)^{1-D}} \Big| a_{k,t,T} - a_k(t/T) \Big| + \Big| a_{l,t,T} - a_l(t/T) \Big|. \end{aligned} \tag{6.38}$$

It now follows completely analogous to the proof of Proposition 2.3 in Dahlhaus and Polonik (2009) that there exists a $0 < \rho < 1$ such that

$$|a_{k,t,T} - a_k(t/T)| \leq C \sum_{i=1}^{k-1} \rho^{k-1} \sum_{j=1}^{p} \Big| a_j(\frac{t-i}{T}) - a_j(\frac{t}{T}) \Big|. \tag{6.39}$$

This yields that the second summand in (6.38) can be bounded by a constant times $l^2 \rho^l / T$ which tends faster to zero than l^{D-1}/T and we therefore restrict ourselves to the first term of (6.38). With (6.39) this term can be bounded through

$$C\sum_{k=0}^{l-1}\frac{1}{(l-k)^{1-D}}\sum_{i=1}^{k-1}\rho^{k-1}\frac{i}{T}\leq C\frac{1}{T}\sum_{k=1}^{l-1}\frac{1}{(l-k)^{1-D}}\rho^{k-1}k^2.$$

By using the fact that $\rho^{k-1}k^2$ is smaller than a constant times k^{-1}, we are left to discuss

$$\frac{1}{T}\sum_{k=1}^{l-1}\frac{1}{(l-k)^{1-D}}\frac{1}{k}.$$

We see from Lemma 6.5 a) that this term is (up to a constant) bounded by $\log(l)l^{D-1}T^{-1}$, which then yields the claim. □

Proof of Theorem 3.1. To show weak convergence we have to prove the following two claims [see van der Vaart and Wellner (1996), Theorem 1.5.4 and 1.5.7]:

(1) Convergence of the finite dimensional distributions

$$\sqrt{T}(\hat{G}_T(y_j)-C_T(y_j))_{j=1,\ldots,K}\xrightarrow{D}(G(y_j))_{j=1,\ldots,K} \tag{6.40}$$

where $y_j=(v_j,\omega_j)\in[0,1]^2$ $(j=1,\ldots,K)$ and $K\in\mathbb{N}$.

(2) Stochastic equicontinuity, i.e.

$$\begin{aligned}&\forall\eta,\varepsilon>0\quad\exists\delta>0:\\&\lim_{T\to\infty}P\Big(\sup_{y_1,y_2\in[0,1]^2:\rho_{2,D}(y_1,y_2)<\delta}\sqrt{T}|(\hat{G}_T(y_1)-C_T(y_1))\\&\qquad\qquad-(\hat{G}_T(y_2)-C_T(y_2))|>\eta\Big)<\varepsilon,\end{aligned}$$

where

$$\rho_{2,D}(y_1,y_2):=\left(\frac{1}{2\pi}\int_0^1\int_0^\pi(\phi_{v_1,\omega_1}(u,\lambda)-\phi_{v_2,\omega_2}(u,\lambda))^2\frac{1}{\lambda^{4D}}d\lambda du\right)^{1/2}$$

with the functions $\phi_{v,\omega}$ defined in (6.2) and $y_i=(v_i,\omega_i)$ for $i=1,2$.

The claim (6.40) can be deduced from Theorem 6.1 a)–c), while stochastic equicontinuity can be concluded along the lines of the corresponding result in Preuß et al. (2012). Note that $\rho_{2,D,T}(y_1,y_2)$ converges to the pseudo distance $\rho_{2,D}(y_1,y_2)$, since $D<1/4$. □

Proof of Lemma 4.3. If we denote with $\psi_{l,p}$ the coefficents in the MA(∞) representation of the process $(1-B)^{-D}Y_t^{AR}(p)$ and with η_l the coefficent which appears if we replace $\hat{D}$ with D in $\hat{\eta}_l$, we obtain with (4.9)

$$\hat{\psi}_{l,p}-\psi_{l,p}=\sum_{k=0}^{l}(\hat{c}_{k,p}\hat{\eta}_{l-k}-c_{k,p}\eta_{l-k}) \tag{6.41}$$

$$= \sum_{k=0}^{l} (\hat{c}_{k,p} - c_{k,p}) \, \hat{\eta}_{l-k} + \sum_{k=0}^{l} c_{k,p} \, (\hat{\eta}_{l-k} - \eta_{l-k}) \, . \tag{6.42}$$

We start with the treatment of the first term and let $l \geq 1$. By employing (4.5) we can apply Cauchy's inequality for holomorphic functions analogously to the proof of Lemma 2.5 in Kreiß et al. (2011) to obtain

$$|\hat{c}_{l,p} - c_{l,p}| = \frac{p^2}{(1+1/p)^l} / \sqrt{T} O_P(1), \quad \text{uniformly in } p, l \in \mathbb{N}.$$

With this bound we get $\sum_{k=0}^{\infty} k^2 |\hat{c}_{k,p} - c_{k,p}| = O_P(p^5/\sqrt{T})$ which directly yields

$$|\hat{c}_{k,p} - c_{k,p}| = O_P(p^5/(\sqrt{T}k^2)), \quad k \neq 0. \tag{6.43}$$

Using (6.37) and properties of the Gamma function we obtain $\hat{\eta}_l \leq C/l^{1-\hat{D}}$, uniformly in $\hat{D}$. Therefore we see with (6.43) that the first term in (6.41) is some $O_P(p^5 T^{-1/2} l^{\hat{D}-1})$. This works again by replacing the sum through its approximating integral (for $k \neq 0, l$) and applying (6.6) as in the proof of Lemma 6.4. Concerning the second summand in (6.41) note that (4.7) implies $c_{k,p} = O_P(1/k^7)$, uniformly in p. If we combine this with (4.5) and the mean value theorem we obtain with standard properties of the gamma function that the second summand in (6.41) is of order $O_P(pT^{-1/2} l^{\max(\hat{D},D)-1})$.

Thus to complete the proof it remains to consider $|\psi_{l,p} - \tilde{\psi}_l|$ which is bounded through

$$\sum_{k=0}^{l} |c_{k,p} - c_k| |\eta_{l-k}|,$$

where c_k are the coefficents in the MA(∞) representation of the process $Y_t = (1-B)^D X_t$, see Assumption 4.1. It follows from (4.2) and Lemma 2.4 in Kreiß et al. (2011) that $|c_{k,p} - c_k| = O(1/(k^2 p^5))$ which implies that $|\psi_{l,p} - \psi_l|$ is of order $O(1/(l^{1-D} p^5))$ as for the first term in (6.41). This yields the claim since $\sqrt{T} = o(p^{10})$. □

Proof of Lemma 4.4. To prove part a), note that $\sup_{(v,\omega) \in [0,1]^2} |\hat{G}_{T,a}(v,\omega)|$ and $\sup_{(v,\omega) \in [0,1]^2} |\hat{G}^*_{T,a}(v,\omega)|$ have the same distribution, because $\tilde{\psi}_l = \psi_l = \psi_l(u)$ for all $u \in [0,1]$ under the null hypothesis and since the Z_t and Z_t^* are both independent and standard normal distributed.

Let us now prove part d) which is essentially a corollary of Theorem 6.1. Before we start, note that when we prove part c) below, we show (6.44). It is thus sufficient for us to prove tightness of $(\beta_T^{1/2} \hat{E}^*_{T,a}(v,\omega))_{(v,\omega) \in [0,1]^2}$ only. To this end, we use again the corresponding result in Preuß et al. (2012) which states that we need a claim such as Theorem 6.1 d). Luckily, $\hat{E}^*_{T,a}(v,\omega)$ is constructed from the stationary process in (4.11), whose coefficients $\tilde{\psi}_l$ satisfy Assumption 2.1.

Therefore Theorem 6.1 holds true with $E_T(v,\omega) = C_T(v,\omega) = 0$. Precisely, we obtain

$$\mathbb{E}|\beta_T^{1/2}(\hat{E}^*_{T,a}(y_1) - \hat{E}^*_{T,a}(y_2))|^k \le (2k)!C^k \rho_{2,T,D}(y_1,y_2)^k(\beta_T/T)^{k/2}$$

for all even integers k. For $D < 1/4$, β_T equals T and $\rho_{2,T,D}(y_1,y_2)$ converges to $\rho_{2,D}(y_1,y_2)$ as before, so we are done. Things change in the other cases. For $D = 1/4$, we have

$$\frac{\beta_T}{T}\rho_{2,T,D}(y_1,y_2)^2 = \frac{1}{T\log N}\sum_{j=1}^{M}\sum_{k=1}^{N/2}(\phi_{v_1,\omega_1,T}(u_j,\lambda_k) - \phi_{v_2,\omega_2,T}(u_j,\lambda_k))^2\frac{1}{\lambda_k},$$

and the limit is notably different from $\rho_{2,D}(y_1,y_2)$ for $D < 1/4$. For example,

$$\begin{aligned}&\frac{1}{T\log N}\sum_{j=1}^{M}\sum_{k=1}^{N/2}\phi^2_{v,\omega,T}(u_j,\lambda_k)\frac{N}{2\pi k}\\ &= \frac{1}{M}\sum_{j=1}^{M}(I_{[0,\frac{\lfloor vM\rfloor}{M}]}(u) - \lfloor vM\rfloor/M)^2\sum_{k=1}^{\lfloor\omega\frac{N}{2}\rfloor}\frac{1}{2\pi k\log N} \to v(1-v)1_{\{\omega\neq 0\}}/(2\pi).\end{aligned}$$

It is simple to show that in general

$$\rho_{2,D}(y_1,y_2) = \begin{cases}0, & \omega_1=\omega_2=0,\\ \frac{v_2(1-v_2)}{(2\pi)^{4D}K_D}, & \omega_1=0,\omega_2\neq 0,\\ \frac{v_1(1-v_1)}{(2\pi)^{4D}K_D}, & \omega_1\neq 0,\omega_2=0,\\ \frac{|v_1-v_2|-(v_1-v_2)^2}{(2\pi)^{4D}K_D}, & \omega_1,\omega_2\neq 0,\end{cases}$$

for $D \ge 1/4$, where $K_{1/4} = 1$ and $K_D = \sum_{k=1}^{\infty} k^{-4D}$ otherwise. Formally, this pseudo distance has to make $[0,1]^2$ a totally bounded space, which is satisfied. Therefore, tightness follows.

We now come to the proof of assertion c). As noted above, it suffices to prove

$$\beta_T^{1/2}\sup_{(v,\omega)\in[0,1]^2}|\hat{E}^*_T(v,\omega) - \hat{E}^*_{T,a}(v,\omega)| = o_P(1), \tag{6.44}$$

and we have

$$\begin{aligned}&\sup_{(v,\omega)\in[0,1]^2}\beta_T^{1/2}|\hat{E}^*_T(v,\omega) - \hat{E}^*_{T,a}(v,\omega)|\\ &\le \beta_T^{1/2}1_{A_T(\alpha)}\sup_{(v,\omega)\in[0,1]^2}|\hat{E}^*_T(v,\omega) - \hat{E}^*_{T,a}(v,\omega)|\\ &\quad + \beta_T^{1/2}1_{A_T^C(\alpha)}\sup_{(v,\omega)\in[0,1]^2}|\hat{E}^*_T(v,\omega) - \hat{E}^*_{T,a}(v,\omega)|.\end{aligned}$$

Note that Lemma 4.3 implies $P(A_T(\alpha)) \to 1$ as $T \to \infty$ for every $\alpha > 0$. Therefore we may restrict our attention to the first term on the right hand side. According to Newey (1991) we have to show the following two claims:

(1) For every $v, \omega \in [0,1]$ we have

$$\beta_T^{1/2} 1_{A_T(\alpha)}(\hat{E}_T^*(v,\omega) - \hat{E}_{T,a}^*(v,\omega)) = o_P(1).$$

(2) For every $\eta, \varepsilon > 0$ there exists a $\delta > 0$ such that

$$\lim_{T\to\infty} P\Big(\sup_{y_1,y_2\in[0,1]^2:\tilde{\rho}(y_1,y_2)<\delta} \beta_T^{1/2} 1_{A_T(\alpha)} |(\hat{E}_T^*(y_1) - \hat{E}_{T,a}^*(y_1)) - (\hat{E}_T^*(y_2) - \hat{E}_{T,a}^*(y_2))| > \eta \Big) < \varepsilon,$$

where $y_i = (v_i, \omega_i)$ for $i = 1, 2$ and $\tilde{\rho}$ is a pseudo distance such that $[0,1]^2$ equipped with $\tilde{\rho}$ is a totally bounded space.

To prove pointwise convergence as in (1) note that

$$\begin{aligned} &\beta_T \mathbb{E}[1_{A_T(\alpha)} |\hat{E}_T^*(v,\omega) - \hat{E}_{T,a}^*(v,\omega)|^2] \\ &= \beta_T \mathbb{E}[1_{A_T(\alpha)}(\hat{E}_T^*(v,\omega) - \hat{E}_{T,a}^*(v,\omega))]^2 + \beta_T \mathrm{Var}\Big((\hat{E}_T^*(v,\omega) - \hat{E}_{T,a}^*(v,\omega)) 1_{A_T(\alpha)} \Big) \end{aligned}$$

Then we use Theorem 6.2 a), b) as well as Assumption 4.2 which yields

$$\beta_T \mathbb{E}[1_{A_T(\alpha)} |\hat{E}_T^*(v,\omega) - \hat{E}_{T,a}^*(v,\omega)|^2] = O\Big(\log(N)^2 N^\alpha p_{min}(T)^{-1} \Big)$$

for all choices of D. Since both $p_{min}(T)$ and N converge to infinity at a polynomial growth in T, convergence to zero follows by choosing α small enough. The same arguments can be used to obtain (2). From Theorem 6.2 c) we have

$$\begin{aligned} &\mathbb{E}|\beta_T^{1/2}((\hat{E}_T^*(y_1) - \hat{E}_{T,a}^*(y_1)) - (\hat{E}_T^*(y_2) - \hat{E}_{T,a}^*(y_2)))|^k \\ &\le (2k)! C^k \tilde{\rho}^k((v_1,\omega_1),(v_2,\omega_2)) (\log(N)^2 N^\alpha p_{min}(T)^{-1})^{k/2}. \end{aligned}$$

We set $a_T = \log(N)^2 N^\alpha p_{min}(T)^{-1}$ and note that it is $\bigcup_{(v,\omega)\in[0,1]^2} \{\phi_{v,\omega,T}\} = \bigcup_{(v,\omega)\in P_T} \{\phi_{v,\omega,T}\}$ with $P_T := \{0, 1/M, 2/M, \dots, 1\} \times \{0, 2/N, 4/N, \dots, 1\}$. We define d_T as the pseudo metric on P_T with respect to which all points have distance a_T and consider the corresponding covering integral of P_T, namely

$$J_T(\delta) = \int_0^\delta \Big[\log(48 N_T^2(x) x^{-1}) \Big]^2 dx,$$

where $N_T(x)$ denotes the covering number of P_T with respect to d_T. The claim can be then deduced completely analogously to the proof of (5.2) in Preuß et al. (2012) if we show that $\lim_{T\to\infty} J_T(\delta)$ converges to zero as $\delta \to 0$. However, for $x < a_T$ it is $N_T(x) = \#P_T = (M+1)(N/2+1) \le 3T$, and for $x \ge a_T$ we get $N_T(x) = 1$. Therefore $J_T(\delta)$ is bounded by

$$\begin{aligned} &\textstyle\int_0^{\min(a_T,\delta)} \log^2(432 T^2 x^{-1}) dx + 1_{\{\delta > a_T\}} \int_{a_T}^\delta \log^2(48 x^{-1}) dx \\ &\le \textstyle\int_0^{a_T} \log^2(432 T^2 x^{-1}) dx + \int_0^\delta \log^2(48 x^{-1}) dx. \end{aligned}$$

Since the second term is independent of T and converges to zero as $\delta \to 0$, we can restrict ourselves to the first integral which equals

$$\int_0^{a_T} \log^2(432T^2)dx + 2\int_0^{a_T} \log(432T^2)\log(x^{-1})dx + \int_0^{a_T} \log^2(x^{-1})dx.$$

For α small enough, this expression obviously converges to zero as $T \to \infty$.

The claim b) finally can be proven in a similar way, but by using Theorem 6.1 instead. For part (1), note that the only difference between $\hat{E}_T(v,\omega)$ and $\hat{E}_{T,a}(v,\omega)$ regards the use of $\psi_l(t/T)$ instead of the true $\psi_{l,t,T}$. Since we are under the null hypothesis, Theorem 6.1 a) shows that the expectation of both $\hat{E}_T(v,\omega)$ and $\hat{E}_{T,a}(v,\omega)$ is the same one, up to an error term $O(1/T)$ which is some $o(\beta_T^{-1/2})$. Similarly, if we compute each of the (co)variances, the first step in the proof of Theorem 6.1 b) is to replace them with the corresponding V_T^{appr} which are all the same quantities. These cancel out, so the remaining step is to prove that the approximation error $O_p(\frac{N}{T^2}(N^{4D-1}\log(N)+1))$ is some $o(\beta_T^{-1})$. This is an immediate consequence of the definition of β_T and the fact that N converges to infinity at a polynomial rate. Therefore, $\beta_T^{1/2}(\hat{E}_T(v,\omega)-\hat{E}_{T,a}(v,\omega))$ converges to zero pointwise as claimed. In order to show (2), note that we are in a position to proceed as in the proof of part d) above. The main difference is, however, that $\hat{E}_T(v,\omega)-\hat{E}_{T,a}(v,\omega)$ is a process whose coefficients $\psi_{t,T,l}-\psi_l(t/T)$ satisfy (2.5). As a consequence,

$$\sup_{t=1,\ldots,T}\Big|\psi_{t,T,l}-\psi_l(t/T)\Big| \le \frac{C}{T}\left(\frac{1}{l^{1-D-\varepsilon}}1_{\{l\neq 0\}} + 1_{\{l=0\}}\right)$$

for some $\varepsilon > 0$ small enough such that $D+\varepsilon < 1/2$. Therefore, using the same proof as for Theorem 6.1 d) we obtain

$$\begin{aligned}&\mathbb{E}|\beta_T^{1/2}((\hat{E}_T(y_1)-\hat{E}_{T,a}(y_1))-(\hat{E}_T(y_2)-\hat{E}_{T,a}(y_2)))|^k\\ &\le (2k)!C^k\rho_{2,T,D+\varepsilon}(y_1,y_2)^k\beta_T^{k/2}T^{-k}.\end{aligned}$$

Tightness follows then in the same way as for part c) above. □

Proof of Theorem 4.5. We begin with the proof of part a). If $D < 1/4$, we know from Theorem 3.1 that $\sup_{(v,\omega)\in[0,1]^2}|\beta_T^{1/2}\hat{E}_T(v,\omega)|$ converges under the null hypothesis in distribution to $\sup_{(v,\omega)\in[0,1]^2}|G(v,\omega)|$, which is a continuous random variable; see Lifshits (1984). Call F_T and F their respective distribution functions.

Now, since F is a continuous distribution function, it is in fact uniformly continuous, and we have furthermore that F_T converges to F uniformly and not just pointwise. Thus, let $\varepsilon > 0$ be arbitrary and choose $\delta > 0$ and $T_1 > 0$ in such a way that for an arbitrary z

$$|F(z)-F(z-\delta)| < \varepsilon,$$

which is possible due to uniform continuity, and that

$$|F(z)-F_T(z)| < \varepsilon \quad\text{and}\quad |F(z-\delta)-F_T(z-\delta)| < \varepsilon$$

for all $T > T_1$, which this time holds due to uniform convergence. Also, we know from Lemma 4.4 and Lemma 6.6 that

$$F_T(Q_T^*(1-\alpha)-\delta)-\varepsilon < 1-\alpha < F_T(Q_T^*(1-\alpha)+\delta)+\varepsilon$$

for any $T \geq \max(T_0, T_1)$. Therefore,

$$\begin{aligned} F_T(Q_T^*(1-\alpha)) \leq & |F_T(Q_T^*(1-\alpha)) - F(Q_T^*(1-\alpha))| \\ & + |F(Q_T^*(1-\alpha)) - F(Q_T^*(1-\alpha)-\delta)| \\ & + |F(Q_T^*(1-\alpha)-\delta) - F_T(Q_T^*(1-\alpha)-\delta)| \\ & + F_T(Q_T^*(1-\alpha)-\delta) \leq 1-\alpha+4\varepsilon \end{aligned}$$

for any $T \geq \max(T_0, T_1)$. In the same way, the lower bound can be obtained. The claim follows since $\varepsilon > 0$ was arbitrary.

In order to obtain the claim in b), we use Lemma 4.4 and Lemma 6.6 to obtain

$$F_T(Q_T^*(1-\alpha)+\delta) \geq 1-\alpha-\varepsilon$$

for any $T \geq T_0$. The same lower bound thus holds for the limes inferior, and as $\varepsilon > 0$ was arbitrary, the desired result follows. □

References

Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, 93:1488–1501. MR1666643

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle*. Budapest, Akademia Kiado, 267–281. MR0483125

Beran, J. (2009). On parameter estimation for locally stationary long-memory processes. *Journal of Statistical Planning and Inference*, 139:900–915. MR2479836

Berg, A., Paparoditis, E., and Politis, D. N. (2010). A bootstrap test for time series linearity. *Journal of Statistical Planning and Inference*, 140:3841–3857. MR2674170

Berkes, I., Horvarth, L., Kokoszka, P., and Shao, Q. M. (2006). On discriminating between long-range dependence and changes in mean. *Annals of Statistics*, 34:1140–1165. MR2278354

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. McGraw Hill, New York. MR0595684

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Verlag, New York. MR1093459

Can, S. U., Mikosch, T., and Samorodnitksy, G. (2010). Weak convergence of the function-indexed integrated periodogram for infinite variance processes. *Bernoulli*, 17(4):995–1015. MR2759166

Chang, C. and Morettin, P. (1999). Estimation of time-varying linear systems. *Statistical Inference for Stochastic Processes*, 2:253–285. MR1919869

CHEN, Y., HÄRDLE, W., and PIGORSCH, U. (2010). Localized realized volatility modeling. *Journal of the American Statistical Association*, 105(492):1376–1393. MR2796557

DAHLHAUS, R. (1988). Empirical spectral processes and their applications to time series analysis. *Stochastic Process and Their Applications*, 30:69–83. MR0968166

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37. MR1429916

DAHLHAUS, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics*, 151:101–112. MR2559818

DAHLHAUS, R. and POLONIK, W. (2006). Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes. *Annals of Statistics*, 34(6):2790–2824. MR2329468

DAHLHAUS, R. and POLONIK, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, 15:1–39. MR2546797

DEHLING, H., ROOCH, A., and TAQQU, M. S. (2013). Nonparametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics*, 40:153–173. MR3024037

DETTE, H., PREUSS, P., and VETTER, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association*, 106(495):1113–1124. MR2894768

DOUKHAN, P., OPPENHEIM, G., and TAQQU, M. S. (2002). *Theory and Applications of Long-Range Dependence*. Birkhäuser, Boston. MR1956041

DWIVEDI, Y. and SUBBA RAO, S. (2010). A test for second order stationarity of a time series based on the discrete fourier transform. *Journal of Time Series Analysis*, 32(1):68–91. MR2790673

EICHLER, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *Journal of Multivariate Analysis*, 99:968–1009. MR2405101

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, 14(2):517–532. MR0840512

FOX, R. and TAQQU, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probability Theory and Related Fields*, 74:213–240. MR0871252

FRYZLEWICZ, P., SAPATINAS, T., and SUBBA RAO, S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93:687–704. MR2261451

GIRAITIS, L., KOUL, H. L., and SURGAILIS, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press. MR2977317

GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1:15–29. MR0605572

HANNAN, E. J. and KAVALIERIS, L. (1986). Regression, autoregression models. *J. Time Ser. Anal.*, 7(1):27–49. MR0832351

Henry, M. and Zaffaroni, P. (2002). The long-range dependence paradigm for macroeconomics and finance. In *Theory and Applications of Long-Range Dependence* (P. Doukhan, G. Oppenheim and M. S. Taqqu). Birkhäuser, Boston, 417–438 MR1957502

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–176. MR0614953

Kokoszka, P. and Mikosch, T. (1997). The integrated periodogram for long-memory processes with finite or infinite variance. *Stochastic Process. Appl.*, 66(1):55–78. MR1431870

Kokoszka, P. S. and Taqqu, M. S. (1995). Fractional ARIMA with stable innovations. *Stochastic Processes and Their Applications*, 60:19–47. MR1362317

Kreiss, J.-P. (1988). *Asymptotic statistical inference for a class of stochastic processes*. Habilitationsschrift, Fachbereich Mathematik, Universität Hamburg.

Kreiss, J.-P. and Paparoditis, E. (2011). Bootstrapping locally stationary processes. Technical report.

Kreiss, J.-P., Paparoditis, E., and Politis, D. N. (2011). On the range of the validity of the autoregressive sieve bootstrap. *Annals of statistics*, 39(4):2103–2130. MR2893863

Lavancier, F., Leipus, R., Philippe, A., and Surgailis, D. (2011). Detection of non-constant long memory parameter. Preprint, Université de Nantes.

Lifshits, M. (1984). Absolute continuity of functionals of "supremum" type for Gaussian processes. *Journal of Soviet Mathematics*, 27:3103–3112.

Mikosch, T. and Starica, C. (2004). Non-stationarities in financial time series, the long range dependence and the IGARCH effects. *The Review of Economics and Statistics*, 86:378–390.

Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and applications to adaptive estimation of evolutionary spectra. *Annals of Statistics*, 25:38–76. MR1429917

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59(4):1161–1167. MR1113551

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. Wiley Series in Probability and Statistics. MR2297359

Palma, W. and Olea, R. (2010). An efficient estimator for locally stationary Gaussian long-memory processes. *Annals of Statistics*, 38(5):2958–2997. MR2722461

Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli*, 15:1190–1221. MR2597589

Paparoditis, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association*, 105(490):839–851. MR2724865

Park, K. and Willinger, W. (2000). *Self-Similar Network Traffic and Performance Evaluation*. Wiley, New York.

Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics*, 23(2):275–290. MR2681201

Preuss, P., Vetter, M., and Dette, H. (2012). A test for stationarity based on empirical processes. *To appear in Bernoulli*.

Roueff, F. and von Sachs, R. (2011). Locally stationary long memory estimation. *Stochastic Processes and Their Applications*, 121:813–844. MR2770908

Sakiyama, K. and Taniguchi, M. (2004). Discriminant analysis for locally stationary processes. *Journal of Multivariate Analysis*, 90:282–300. MR2081780

Sibbertsen, P. and Kruse, R. (2009). Testing for a change in persistence under long-range dependencies. *Journal of Time Series Analysis*, 30:263–285. MR2512240

Starica, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87:503–522.

Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Annals of Statistics*, 36(4):1879–1924. MR2435459

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, Berlin. MR1385671

von Sachs, R. and Neumann, M. H. (2000). A wavelet-based test for stationarity. *Journal of Time Series Analysis*, 21:597–613. MR1794489

Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. HUppsala: Almqvist and Wiksell. MR0040634

Zygmund, A. (1959). *Trigonometric Series I*. Cambridge University Press. MR0107776