How much information does dependence between wavelet coefficients contain?

Carsten Jentsch[*] Claudia Kirch[†]

September 8, 2015

Abstract Motivated by several papers which propose statistical inference assuming independence of wavelet coefficients for both short- as well as long-range dependent time series, we focus exemplary on the sample variance and investigate the influence of the dependence between wavelet coefficients to this statistic. To this end, we derive asymptotic distributional properties of the sample variance for a time series which is synthesized ignoring some or all dependence between wavelet coefficients. We show that the second order properties differ from the ones of the true time series whose wavelet coefficients have the same marginal distribution except in the independent Gaussian case. This holds true even if the dependency is correct within each level and only the dependence between levels is ignored. For the example of sample autocovariances and sample autocorrelations at lag one, we indicate that already first order properties are erroneous in these cases. In a second step, several non-parametric bootstrap schemes in the wavelet domain are investigated which take more and more dependence into account until finally the full dependency is mimicked. We obtain very similar results, where only a bootstrap mimicking the full covariance structure correctly can be valid asymptotically. A simulation study supports our theoretical findings for the wavelet domain bootstraps. For long-range dependent time series with long-memory parameter $d > 1/4$, we show that some additional problems occur which cannot be solved easily without using additional information for the bootstrap.

Keywords: bootstrap inconsistency, decorrelation, discrete wavelet transform, long-range dependence, synthesis of time series, wavestrapping.

Subject classification: 62G09, 62M10, 42C40

[*]University of Mannheim, Department of Economics, L7, 3-5, 68131 Mannheim, Germany; cjentsch@mail.uni-mannheim.de

[†]Otto von Guericke University Magdeburg, Institute for Mathematical Stochastics, Postfach 4120, D–39016 Magdeburg, Germany; claudia.kirch@ovgu.de

1. Introduction

In the last 20-30 years wavelet analysis has become a useful and popular tool for statistical data analysis, where its advantages over the classical Fourier analysis lies in the fact that it is local in time and frequency. The books by Nason (2010), Percival and Walden (2000) and Vidakovic (2009) give a nice overview over current applications in statistics, where the book by Percival and Walden (2000) focuses on time series applications. While most applications coming from the area of signal processing have been treated extensively in the engineering literature, there has been some interest in the statistical properties of wavelet coefficients of time series.

This paper focuses on the question how much information the dependence between wavelet coefficients contains. More precisely, we investigate the effect of neglecting this dependence on the asymptotic variance of a given statistic. The results help to understand the asymptotic behavior of related wavelet based resampling techniques. Finally, we investigate the potential of wavelet based resampling methods for a long-memory parameter $d > 1/4$. All three results urge to be extremely careful with inferential results based on the independence assumption of wavelet coefficients within as well as between scales.

Wavelet methods are related to Fourier analysis, where the Fourier coefficients have the nice property of being asymptotically independent (but not identically distributed), a fact, that can be exploited in the estimation of such fundamental time series characteristics as the spectral density or autocorrelations. It has also given rise to a multitude of useful frequency domain bootstrap method, for recent reviews we refer to Paparoditis (2002), Kreiss and Paparoditis (2011) and Kreiss and Lahiri (2012).

The corresponding statistical inference is related to using the Whittle likelihood (Whittle (1957)) for statistical inference of time series, which has not only been successfully applied to Frequentist but also to Bayesian statistical inference (see e.g. Carter and Kohn (1997)). However, as shown by Contreras-Cristán et al. (2006) the loss of efficiency can be substantial.

Similarly, wavelet transforms also frequently exhibit a whitening effect leading to less correlated wavelet coefficients even for long-range dependent time series whose long-memory covariance structure in the time domain is very dense (see e.g. Dijkerman and Mazumdar (1994), Craigmile and Percival (2005), Kaplan and Kuo (1993), Tewfik and Kim (1992), Fan (2003), Mielniczuk and Wojdyllo (2007) or Theorem 9.2.2 in Chapter 9.5.3 in Vidakovic (2009) as well as Proposition 13.1 in Walter (1994)). This whitening effect has first been exploited by Wornell and Oppenheim (1992) for the analysis of $1/f$-processes and later been pursued by others (confer e.g. Jensen (2000) or Craigmile et al. (2005)). Flandrin (1992) and McCoy and Walden (1996) additionally propose to make use of this effect for the synthesis of stationary long-memory processes.

Moulines et al. (2007a,b, 2008) and Faÿ et al. (2009) propose to use a pseudo-maximum-likelihood estimator based on the independence assumption between wavelet coefficients to estimate the long-memory parameter in long-range dependent time series. They provide rigorous asymptotic theory under misspecification of this likelihood for a large class of time domain models including Gaussian linear processes. The aim of this paper is to shed light on the asymptotic properties of wavelet based resampling techniques, which have been proposed but only empirically been studied in the literature. Percival et al.

(2000) adopt the term wavestrapping to use the whitening effect of the wavelet transform to bootstrap (long-memory) time series, where the main advantage is that the long-memory parameter d does not need to be known or estimated to obtain bootstrap time series in the time domain. A related approach has been proposed by Feng et al. (2005) using an AR-bootstrap in the wavelet domain to incorporate some of the dependence between wavelet coefficients. Sabatini (1999) proposes to use a level-wise moving block bootstrap for the construction of confidence intervals, while Angelini et al. (2005) propose a levelwise stationary bootstrap for hierarchical processes. Different versions of block-type bootstrap schemes have been applied to wavelet leaders instead of wavelet coefficients for multifractal analysis in Wendt et al. (2007), Wendt and Abry (2007) and Wendt et al. (2009).

Bullmore et al. (2001), Breakspear et al. (2003) and Whitcher (2006) use a related wavelet-based bootstrap for the analysis of functional magnetic resonance imagining (fMRI) experiments. Ko and Vannucci (2006) propose to use an approximate likelihood based on uncorrelated wavelet coefficients for Bayesian inference of long-memory time series and have recently applied this approach to fMRI data (confer Jeong et al. (2013)). On the other hand, empirical evidence by Aston et al. (2005) indicate that independence between wavelet coefficients cannot be assumed for fMRI data.

In this paper, we aim to shed light on the question how much information is actually coded in the dependence between wavelet coefficients by analyzing its asymptotic effect on autocovariances and autocorrelations, where we focus on the statistic of the sample variance. This is the key to understanding the asymptotics of wavelet based resampling methods. We choose the sample variance for that purpose because its analysis in the wavelet domain has been addressed extensively in the literature (cf. e.g. Percival and Mondal (2012)) and it is the simplest statistic for which this analysis can be applied. The sample mean is coded in the final scaling coefficient, so that the information in the wavelet coefficients do not help to explain its distributional properties. This is similar to Fourier analysis, where the sample mean is also coded as the coefficient at frequency 0 and needs to be treated differently from the other frequencies; see Proposition 10.3.1 in Brockwell and Davis (1991).

In Section 2, we introduce the discrete wavelet transform (DWT) and provide most of the notation used throughout the paper. In Section 2.1, we compare the asymptotic distribution of the sample variance for a given time series with the asymptotic distribution for a (synthesized) time series with wavelet coefficients that have the same marginal distribution but ignore the dependency between coefficients. It turns out that the two distributions differ for the case of a simple 1-dependent (i.e. Moving-Average) Gaussian time series, but also for autoregressive Gaussian time series. In the non-Gaussian case, we find that they differ already for i.i.d. data. In a next step, we investigate in Section 2.2 the situation where the dependence between coefficients at the same level is mimicked correctly and merely the dependence between levels is ignored. This is of interest, because it is sometimes argued in the literature that the dependence across levels is smaller and can therefore be ignored more easily; see e.g. Sabatini (1999) and Angelini et al. (2005) and the analysis of AR(1) data based on mutual dependence between interscale and intrascale wavelet coefficients by Liu and Moulin (2000). Unfortunately, while this improves the situation to some extent, it still yields asymptotically inconsistent results. Somewhat surprisingly, the influence of the dependence between levels seems to be even larger than within levels at least for the Haar wavelet basis. In Section 2.3, we address sample autocovariances and autocorrelations at lag one and prove that already their first order asymptotics are erratic for the corresponding synthesized time series.

In Section 3 we consider several nonparametric bootstrap procedures in the wavelet domain, where more and more of the dependency structure is taken into account. We analyse distributional properties of the corresponding bootstrap sample variance, where it becomes apparent that the same restrictions as in Section 2 apply. In the last step, we propose a new bootstrap scheme, which does have the potential to yield consistent results but at the expense of being very complicated, which seems inappropriate in a situation, where a standard block bootstrap yields consistent results and has good small sample properties. We illustrate our theoretical findings by some simulations. Because the block bootstrap can no longer work for long-range dependent time series with long-memory parameter $d > 1/4$, we shed some light on the validity of wavelet domain resampling methods in that situation in Section 4. It turns out that even the single wavelet coefficient of the coarsest scale has a non-negligible influence on the asymptotic distribution of the sample variance. Because this coefficient can never be mimicked correctly by standard resampling methods simply because there is no duplicated information as n increases, all such methods fail. A possible way around are subsampling methods if the long-memory parameter d is known but then this methodology could immediately be applied to the statistic of interest in the time domain; see Chapter 10 in Lahiri (2003) and Beran et al. (2013), Chapter 10. Finally, Section 5 gives some conclusions, while the proofs are given in a supplement file.

In addition to our theoretical findings and our own small simulation study, there are already several extensive simulation studies in the literature supporting our theoretical findings (cf. Sabatini (1999); Percival and Walden (2000); Angelini et al. (2005); Feng et al. (2005); Tang et al. (2008)). Furthermore, the data analysis in Aston et al. (2005) indicates that the independence assumption is not justified for fMRI data and can lead to large statistical errors.

2. The influence of dependence between wavelet coefficients

In this section we quantify the impact of dependence between wavelet coefficients by focusing on the asymptotic distribution of the sample variance of a real-valued stationary time series in comparison to the sample variance of a synthesized time series, where the marginal distributions of the coefficients are the same but all coefficients respectively coefficients of different scales are independent. Additionally, sample autocovariances/autocorrelations at lag one are addressed in Section 2.3. To simplify the considerations, we assume that $X_1, \ldots, X_n$ is a stationary linear time series with mean zero and consider the sample variance

$$T_n(\boldsymbol{X}) = \frac{1}{n}\sum_{t=1}^{n}\left(X_t - \bar{X}_n\right)^2 = \frac{1}{n}\sum_{t=1}^{n}X_t^2 - \bar{X}_n^2, \quad \bar{X}_n = \frac{1}{n}\sum_{t=1}^{n}X_t. \tag{2.1}$$

In order to understand the effect of ignoring dependence between wavelet coefficients on the asymptotic distribution of the above sample variance in the dependent case, it is useful to remember the asymptotic distributions as given in the following two theorems. First, in Theorem 2.1, the short-range dependent (SRD) case is considered, where we explicitly give the formula for the asymptotic variance for some important examples that will be considered in detail below. This result can be found in Proposition 7.3.4 in Brockwell and Davis (1991).

Theorem 2.1 (Asymptotics for the sample variance under SRD).

(i) Let $\{X_t, t \in \mathbb{Z}\}$ be a linear process with $X_t = \sum_{\nu=-\infty}^{\infty} \alpha_\nu \epsilon_{t-\nu}$, where $\{\epsilon_t, t \in \mathbb{Z}\}$ with $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{E}(\epsilon_t^2) = \sigma^2 \in (0, \infty)$, $\mathrm{E}(\epsilon_t^4) = \eta\sigma^4 \in [0, \infty)$ is an i.i.d. white noise process, $\sum_{\nu=-\infty}^{\infty} |\alpha_\nu| < \infty$ and $\gamma_{\boldsymbol{X}}(h) = \mathrm{E}(X_{t+h}X_t)$. Then, it holds $\sqrt{n}\,(T_n(\boldsymbol{X}) - \gamma_{\boldsymbol{X}}(0)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V)$ with

$$V = \gamma_{\boldsymbol{X}}^2(0)\,(\eta - 3) + 2\sum_{h=-\infty}^{\infty} \gamma_{\boldsymbol{X}}^2(h), \tag{2.2}$$

which simplifies to $V = 2\sum_{h=-\infty}^{\infty} \gamma_{\boldsymbol{X}}^2(h)$ if $\{X_t, t \in \mathbb{Z}\}$ is Gaussian.

(ii) For the i.i.d. situation with $X_t = \epsilon_t$, i.e. $\alpha_\nu = 0$ for all $\nu \neq 0$, the variance in (2.2) simplifies to

$$V = \sigma^4(\eta - 3) + 2\sigma^4. \tag{2.3}$$

(iii) If $\{X_t, t \in \mathbb{Z}\}$ is Gaussian and 1-dependent, i.e. $\gamma_{\boldsymbol{X}}(h) = 0$ for all $|h| > 1$, we have $X_t = \alpha_1\epsilon_{t-1} + \epsilon_t$ and the variance in (2.2) simplifies to

$$V = 2\gamma_{\boldsymbol{X}}^2(0) + 4\gamma_{\boldsymbol{X}}^2(1). \tag{2.4}$$

As we treat also long-range dependent (LRD) time series in Sections 3 and 4 below, we provide in Theorem 2.2 the correct asymptotic distributions of the sample variance for FARIMA(p,d,q)-processes, which constitute an important example of LRD time series with long-memory parameter d. These results can be found in Hosking (1996), Theorems 3–5.

Theorem 2.2 (Asymptotics for the sample variance under LRD). *Let $\{X_t, t \in \mathbb{Z}\}$ be a linear process as in Theorem 2.1 (i) but with $\sum_{\nu=-\infty}^{\infty} |a_\nu| = \infty$ such that $\gamma_{\boldsymbol{X}}(r) = O(|r|^{2d-1})$ as $|r| \to \infty$ with $0 < d < 1/2$. Then, we have*

$$a(n, d)(T_n(\mathbf{X}) - \gamma_{\boldsymbol{X}}(0)) \xrightarrow{\mathcal{D}} Z_d, \tag{2.5}$$

where Z_d is a nondegenerate random variable, which is Gaussian for $0 < d \leqslant 1/4$ and Rosenblatt for $1/4 < d < 1/2$ and

$$a(n, d) = \begin{cases} \sqrt{n}, & 0 \leqslant d < 1/4, \\ \sqrt{n/\log n}, & d = 1/4, \\ n^{1-2d}, & 1/4 < d < 1/2. \end{cases} \tag{2.6}$$

Let $\boldsymbol{O}$ be an orthonormal transformation, i.e. $\boldsymbol{O}^T\boldsymbol{O} = \boldsymbol{O}\boldsymbol{O}^T = \mathbf{Id}$, where $\mathbf{Id}$ denotes the identity matrix. Then, we define the coefficients corresponding to the transformation of $\boldsymbol{X}$ by $\boldsymbol{C_X} = (c_{\boldsymbol{X}}(1), \ldots, c_{\boldsymbol{X}}(n))^T$, that is, $\boldsymbol{C_X} = \boldsymbol{O}\boldsymbol{X}$. With the exception of Theorem 2.3 we will focus on wavelet transforms $\boldsymbol{W}$ and additionally assume for simplicity that $n = 2^D$ for some $D \in \mathbb{N}$. In time series analysis, where only discrete time points are observed, discrete versions of wavelets (similar to the discrete Fourier transform) are used. We will follow the notation (and ordering) of Chapter 4 in Percival and Walden (2000). Such wavelet bases are constructed from a wavelet filter $\{h_j : j = 0, \ldots, M-1\}$ with some given width $M = 2m \leqslant n$ for some integer m fulfilling

$$\sum_{j=0}^{M-1} h_j = 0, \qquad \sum_{j=0}^{M-1} h_j^2 = 1, \qquad \sum_{j=0}^{M-1} h_j h_{j+2k} = 0 \quad \text{for all } k \in \mathbb{Z} \setminus \{0\},$$
$$h_0 \neq 0, \quad h_{M-1} \neq 0.$$

The first $n/2$ rows $\boldsymbol{h}_{1,i}$, $1 \leqslant i \leqslant n/2$, of the orthogonal matrix $\boldsymbol{W} = (W(i,j))_{i,j=1,\ldots,n}$ are obtained by circularly shifting the (row) vector $\boldsymbol{h} = (h_1, h_0, 0, \ldots, 0, h_{M-1}, \ldots, h_2)$ by $2(i-1)$ to the right, i.e.

$$\boldsymbol{h}_{1,i} := W(i,\cdot) = \boldsymbol{h}\boldsymbol{S}^{2(i-1)}, \quad i = 1, \ldots, n/2, \tag{2.7}$$

where

$$\boldsymbol{S} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

The next $n/4$ matrix rows are calculated by means of the Pyramid algorithm, which uses a 'quadrature mirror' filter $g_l = (-1)^l g_{M-1-l}$ in a similar way as above to obtain so called scaling coefficients, which are then transformed to wavelet coefficients by applying the original filter $\boldsymbol{h}$ (and its evenly shifted versions) to those scaling coefficients. Details can be found in the book by Percival and Walden (2000), Chapter 4.4 – 4.6. Consequently, the original time series is filtered with $\boldsymbol{h}_{2,1}$, which has $3M-2$ (circularly consecutive) elements, where at least the first and last are non-zero, while all other elements of the vector are zero, i.e.

$$\boldsymbol{h}_{2,1} = (h_3^{(2)}, \ldots, h_0^{(2)}, 0, \ldots, 0, h_{3M-3}^{(2)}, \ldots, h_4^{(2)}) \tag{2.8}$$

with $h_0^{(2)} \neq 0$ as well as $h_{3M-1}^{(2)} \neq 0$ (only depending on $h_0, \ldots, h_{M-1}$). The next $n/4-1$ rows oft the wavelet matrix are obtained by shifting $\boldsymbol{h}_{2,1}$ successively by 4 elements to the right (in a circular way), i.e. $\boldsymbol{h}_{2,i} = \boldsymbol{h}_{2,1}\,\boldsymbol{S}^{4(i-1)}$, $i = 1, \ldots, n/4$. As $n = 2^D$, this procedure can be continued until only one scaling coefficient remains which is given by $1/\sqrt{n}\,(1, \ldots, 1)$ constituting the last row of the matrix $\boldsymbol{W}$ (confer Exercise 97 in Percival and Walden (2000)). At scale k we obtain (confer page 152 in Percival and Walden (2000))

$$\boldsymbol{h}_{k,1} = (h_{2^k-1}^{(k)}, \ldots, h_0^{(k)}, 0, \ldots, 0, h_{M_k-1}^{(k)}, \ldots, h_{2^k}^{(k)}), \tag{2.9}$$

with $M_k = (2^k-1)(M-1)+1$, which is the number of circularly consecutive (potentially) non-zero elements of $\boldsymbol{h}_{k,1}$. The next $n/2^k - 1$ elements are obtained by shifting $\boldsymbol{h}_{k,1}$ successively by 2^k elements to the right (in a circular way), i.e. $\boldsymbol{h}_{k,i} := \boldsymbol{h}_{k,1}\boldsymbol{S}^{2^k(i-1)}$, $i = 1, \ldots, n/2^k$. In particular, we get for $l < \lceil \frac{M_k}{2^k} \rceil$

$$d_k(l) = \sum_{j=0}^{l2^k-1} h_j^{(k)} X_{l2^k-j} + \sum_{j=l2^k}^{M_k-1} h_j^{(k)} X_{n+l2^k-j}, \tag{2.10}$$

as well as for $l \geqslant \lceil \frac{M_k}{2^k} \rceil$, i.e. if no circular wrapping is present,

$$d_k(l) = \sum_{j=0}^{M_k-1} h_j^{(k)} X_{l2^k-j}. \tag{2.11}$$

Often, the wavelet transform is stopped at a certain point and the scaling coefficients no longer split into more wavelet coefficients. If this approach is taken in the following analysis, we reach the same statistical conclusions as the arguments in the remainder of this section are (with the exception of the explicit formulas for the Haar wavelet) only based on the first and second level of wavelet coefficients (see also Remarks 2.1 and 2.3). In conclusion, the matrix $\boldsymbol{W}$ is given by

$$\begin{aligned} &W\left(n\left(1-(1/2)^{k-1}\right)+j,\cdot\right) = \boldsymbol{h}_{k,j}, \quad k=1,\ldots,D, \quad j=1,\ldots,n_k,\\ &W(n,\cdot) = \frac{1}{\sqrt{n}}(1,\ldots,1) = \frac{1}{\sqrt{n}}\mathbf{1}^T, \end{aligned} \tag{2.12}$$

where we set $n_k = n/2^k$. Correspondingly, we obtain the following wavelet coefficients (plus the final scaling coefficient)

$$\boldsymbol{D}_{\boldsymbol{X}} = \boldsymbol{W}\boldsymbol{X} = (d_{\boldsymbol{X},1}(1),\ldots,d_{\boldsymbol{X},1}(n/2),d_{\boldsymbol{X},2}(1),\ldots,d_{\boldsymbol{X},2}(n/4),\ldots,d_{\boldsymbol{X},D}(1),v_{\boldsymbol{X},D}(1))^T \tag{2.13}$$

with $v_{\boldsymbol{X},D}(1) = \sqrt{n}\,\bar{X}_n$. In this notation, $d_{\boldsymbol{X},k}(l)$ is the lth coefficient belonging to the kth level.

The simplest discrete wavelets, the so called Haar wavelets, are obtained from $h_0 = -1/\sqrt{2}$ and $h_1 = 1/\sqrt{2}$ with $M = 2$. In this case, the general formula

$$h_{k,j}(p) = 2^{-k/2}\left(1_{\{(j-1)2^k < p \leqslant (j-1)2^k + 2^{k-1}\}} - 1_{\{(j-1)2^k+2^{k-1} < p \leqslant j2^k\}}\right) \tag{2.14}$$

holds. More explicitly, for the first level $k = 1$, we get

$$\boldsymbol{h}_{1,1} = \frac{1}{\sqrt{2}}(1,-1,0,\ldots,0), \quad \boldsymbol{h}_{1,2} = \frac{1}{\sqrt{2}}(0,0,1,-1,0,\ldots,0), \quad \ldots \quad ,\boldsymbol{h}_{1,n/2} = \frac{1}{\sqrt{2}}(0,\ldots,0,1,-1),$$

for $k = 2$,

$$\begin{aligned} &\boldsymbol{h}_{2,1} = \frac{1}{2}(1,1,-1,-1,0,\ldots,0), \quad \boldsymbol{h}_{2,2} = \frac{1}{2}(0,0,0,0,1,1,-1,-1,0,\ldots,0),\\ &\ldots \quad ,\boldsymbol{h}_{2,n/4} = \frac{1}{2}(0,\ldots,0,1,1,-1,-1) \end{aligned}$$

etc. for $k \geq 3$. That is, $\boldsymbol{h}_{k,1}$ is given by a vector with first 2^{k-1} times 1 followed by 2^{k-1} times -1 and zero elsewhere, scaled by $1/2^{k/2}$. The vector $\boldsymbol{h}_{k,j}$ is obtained by shifting $\boldsymbol{h}_{k,1}$ by $(j-1)2^k$ to the right. The Haar wavelets are the simplest wavelets from the class of Daubechies wavelets which were introduced by Daubechies (1992) by imposing certain regularity conditions based on vanishing moments; compare e.g. Percival and Walden (2000), Section 11.9.

2.1. Ignoring dependence between wavelet coefficients

In this section we analyze how much information is lost by ignoring all dependence between coefficients. To quantify this, we consider the following synthesized time series with independent coefficients in the wavelet (or more general transformation) domain, which have the same marginal distributions as the coefficients belonging to the original time series $\{X_t\}$:

$$\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T = \boldsymbol{O}^T \boldsymbol{C}_{\boldsymbol{Y}}, \qquad \boldsymbol{C}_{\boldsymbol{Y}} = (c_{\boldsymbol{Y}}(1), \ldots, c_{\boldsymbol{Y}}(n))^T, \tag{2.15}$$

where $\{c_{\boldsymbol{Y}}(1), \ldots, c_{\boldsymbol{Y}}(n)\}$ are independent with $c_{\boldsymbol{Y}}(j) \stackrel{\mathcal{D}}{=} c_{\boldsymbol{X}}(j)$. It is important to note that the distribution of $\boldsymbol{Y}$ depends on the distribution of $\boldsymbol{X}$ but ignores all dependence between the coefficients. If $\boldsymbol{O} = \boldsymbol{W}$ is a wavelet transform, we write $\boldsymbol{Y} = \boldsymbol{W}^T \boldsymbol{D}_{\boldsymbol{Y}}$ with

$$\boldsymbol{D}_{\boldsymbol{Y}} = (d_{\boldsymbol{Y},1}(1), \ldots, d_{\boldsymbol{Y},1}(n_1), d_{\boldsymbol{Y},2}(1), \ldots, d_{\boldsymbol{Y},2}(n_2), \ldots, d_{\boldsymbol{Y},D}(1), v_{\boldsymbol{Y},D}(1))^T, \tag{2.16}$$

where $\{d_{\boldsymbol{Y},k}(l), k = 1, \ldots, D,\ l = 1, \ldots, n_k = n/2^k\}$ are independent with $d_{\boldsymbol{Y},k}(l) \stackrel{\mathcal{D}}{=} d_{\boldsymbol{X},k}(l)$. The sample variance $T_n(\cdot)$ can be written based only on the wavelet coefficients, hence does not depend on the scaling coefficient (confer Lemma A.1).

This setup has explicitely been assumed by Craigmile et al. (2005) in order to develop a rigorous and valid (asymptotic) theory. However, as the below theorem indicates, the corresponding statistical inference such as confidence bounds have to be handled with care as the dependence between coefficients is not in general asymptotically negligible (not even for Gaussian data) and may lead to a systematic error. Nevertheless, in many situations this model error may be smaller than an estimation error associated with a truly nonparametric procedure or nonparametric inference may not be feasible at all.

The following theorem gives the asymptotic distribution of the sample variance as defined in (2.1) of the synthesized time series $\{Y_t\}$ based on an i.i.d. sample $\{X_t\}$.

Theorem 2.3 (i.i.d. case)**.** *Let $\{X_t\}$ be i.i.d. with $\mathrm{E}\, X_t = 0$ and $\mathrm{E}\, |X_t|^4 < \infty$ and denote by $\{Y_t\}$ the corresponding synthesized time series as given in* (2.15) *ignoring all dependence between coefficients.*

(a) Then, it holds $\mathrm{E}\, T_n(\boldsymbol{Y}) = \mathrm{E}\, T_n(\boldsymbol{X})$, but

$$n \operatorname{var} T_n(\boldsymbol{Y}) = \lambda_{\boldsymbol{O}_n} \sigma^4 (\eta - 3) + 2\sigma^4 + O\left(\frac{1}{n}\right),$$

where $0 < \lambda_{\boldsymbol{O}_n} \leqslant 1$. Furthermore, we get the following assertions about $\lambda_{\boldsymbol{O}_n}$:

(i) It holds $\lambda_{\boldsymbol{O}_n} = 1$ iff $\boldsymbol{O}_n$ is a permutation matrix, i.e. has exactly one 1 in every row and column and all other entries are zeros.

(ii) If $\boldsymbol{O}_n = \boldsymbol{W}_n$ is a discrete wavelet matrix based on the filter $h_0, \ldots, h_{M-1}$, then we get the exact representation

$$n \operatorname{var} T_n(\boldsymbol{Y}) = \lambda_{\boldsymbol{W}_n} \sigma^4 (\eta - 3) + \left(2 - \frac{2}{n}\right) \sigma^4,$$

where $\lambda_{\boldsymbol{W}_n}$ is bounded away from one, i.e.

$$\lambda_{\boldsymbol{W}_n} \leqslant \frac{1}{2} + \frac{1}{2}\left(\sum_{j=0}^{M-1} h_j^4\right) < 1.$$

(iii) If $\boldsymbol{O}_n = \boldsymbol{W}_{n,H}$ is the Haar-wavelet transform, then $\lambda_{\boldsymbol{W}_{n,H}} \to 1/3$.

(b) If additionally $\mathrm{E}\,|X_t|^6 < \infty$ and X_t^2 is non-degenerate, then

$$\frac{T_n(\boldsymbol{Y}) - \mathrm{E}\,T_n(\boldsymbol{Y})}{\sqrt{\operatorname{var} T_n(\boldsymbol{Y})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

In fact, for permutation matrices as in (a)(i) due to the exchangeability of the original time series the synthesized time series does not only mimic the marginal distribution of the coefficients correctly but the full joint distribution. In all other cases while the coefficients are uncorrelated by definition the orthonormal transformation does induce some dependence between coefficients that is not negligible asymptotically.

The above theorem states in particular that the asymptotic distribution is only the same for both the original as well as synthesized time series in the i.i.d. case if either $\eta = 3$ as for example in the Gaussian case or the transformation is in some asymptotic sense given by a permutation matrix. However, the latter can never happen for a wavelet transform. This observation is summarized in the following corollary:

Corollary 2.4. *Let $d_2(\cdot,\cdot)$ denote the Mallows' metric. Then, under the assumptions of Theorem 2.3 (b) and for a wavelet transform, i.e. $\boldsymbol{O}_n = \boldsymbol{W}_n$, based on a given filter (or more generally if $\limsup_{n\to\infty} \lambda_{\boldsymbol{O}_n} < 1$), then $\mathrm{E}\,T_n(\boldsymbol{Y}) = \mathrm{E}\,T_n(\boldsymbol{X})$, but*

$$d_2\left(\sqrt{n}(T_n(\boldsymbol{Y}) - \mathrm{E}\,T_n(\boldsymbol{Y}))\,,\, \sqrt{n}(T_n(\boldsymbol{X}) - \mathrm{E}\,T_n(\boldsymbol{X}))\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \Leftrightarrow \quad \eta = 3.$$

While the above theorem and corollary do already show restrictions on how synthesized (and in later chapters bootstrapped) time series can be used to obtain statistically valid assertions in the non-Gaussian case, there may still be useful applications in particular for Gaussian data. In particular, for frequency domain methods this has turned out to be the case. A detailed discussion and comparison can be found in Section 2.4.1.

For this reason, we will now restrict our considerations to the Gaussian case and focus on wavelet bases as in (2.12). While in this case the synthesized time series gives consistent results for i.i.d. data this is no longer the case if dependence is present. More precisely, we show that even the sample variance is not correctly mimicked if the dependence between wavelet coefficients is not taken into account. In Theorem 2.5, we show that this is already true for a 1-dependent time series despite its relatively mild dependency structure. Similar arguments can be employed if stronger dependencies are present and Theorem 2.7 addresses the case of a Gaussian AR(1) time series.

To make the above discussion more precise, we will compare the variances of $T_n(\boldsymbol{X})$ and $T_n(\boldsymbol{Y})$ for a Gaussian and 1-dependent time series in the following theorem. For the variance of $T_n(\boldsymbol{X})$, we get the exact representation

$$n\operatorname{var}(T_n(\boldsymbol{X})) = \left(2 - \frac{2}{n}\right)\gamma_{\boldsymbol{X}}^2(0) + \alpha_n \gamma_{\boldsymbol{X}}(0)\gamma_{\boldsymbol{X}}(1) + \beta_n \gamma_{\boldsymbol{X}}^2(1), \tag{2.17}$$

where $\alpha_n \to 0$ and $\beta_n \to 4$ as $n \to \infty$; compare Theorem 2.1(iii).

Theorem 2.5 (1-dependent case)**.** *Suppose $\{X_t, t \in \mathbb{Z}\}$ is Gaussian and 1-dependent with $\mathrm{E}\, X_t = 0$ and $\mathrm{E}\,|X_t|^4 < \infty$ and denote by $\{Y_t\}$ the corresponding synthesized time series ignoring all dependence between coefficients as given in* (2.15) *and* (2.16) *using a discrete wavelet transform based on the filter $h_0, \ldots, h_{M-1}$.*

(a) Then, $\mathrm{E}\, T_n(\boldsymbol{Y}) = \mathrm{E}\, T_n(\boldsymbol{X})$, but there exists a constant $\beta_{\boldsymbol{W}_n}$ only depending on $\boldsymbol{W}_n$ such that

$$n \operatorname{var}(T_n(\boldsymbol{Y})) = \left(2 - \frac{2}{n}\right) \gamma_{\boldsymbol{X}}^2(0) + \alpha_n \gamma_{\boldsymbol{X}}(0)\gamma_{\boldsymbol{X}}(1) + \beta_{\boldsymbol{W}_n} \gamma_{\boldsymbol{X}}^2(1),$$

with $\beta_n - \beta_{\boldsymbol{W}_n} \geq (h_0 h_{M-1})^2 > 0$ for all $n \geq 2M$.

(b) If $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$ is the Haar-wavelet transform, then $\beta_{\boldsymbol{W}_{n,H}} \to 16/7$.

The above theorem shows that ignoring the dependence between coefficients does not lead to asymptotically consistent results for any fixed wavelet basis not even in the Gaussian case, because the asymptotic variance of $\sqrt{n}\, T_n(\boldsymbol{X})$ requires a factor 4 in front of $\gamma_{\boldsymbol{X}}^2(1)$ (compare with Theorem 2.1 (iii) and (2.17)). This is different from Fourier analysis which consistently mimics the sample variance in the Gaussian SRD case if the dependence (but not heteroscedasticity) between Fourier coefficients is ignored.

The given lower bound for the error in (a) is very conservative and one can expect the true error to be much larger as can e.g. be seen for the Haar basis when comparing it with the true (asymptotic) error.

The main consequence of the above theorem is summarized in the following corollary.

Corollary 2.6. *Let the assumptions of Theorem 2.5 (a) hold and $\gamma_{\boldsymbol{X}}(1) \neq 0$. Then, $\mathrm{E}\, T_n(\boldsymbol{Y}) = \mathrm{E}\, T_n(\boldsymbol{X})$, but*

$$d_2\left(\sqrt{n}(T_n(\boldsymbol{Y}) - \mathrm{E}\, T_n(\boldsymbol{Y}))\,,\, \sqrt{n}(T_n(\boldsymbol{X}) - \mathrm{E}\, T_n(\boldsymbol{X}))\right) \underset{n\to\infty}{\not\to} 0.$$

Remark 2.1. *The assertions of the above theorem and the corollary remain true if we synthesize a time series using independent wavelet coefficients of levels 1 to j ($j \geqslant 1$), which are also independent of the remaining scaling coefficients, but mimic the joint distribution of the scaling coefficients correctly. The reason is that the proof is merely based on the finest level ($k = 1$).*

The following theorem shows that similarly to the one-dependent case the asymptotic variances of the sample variances differ for the synthesized and for the true time series in the case of an autoregressive time series of order 1 even though the decorrelation property of wavelets is more effective for AR(1) than for 1-dependent time series (cf. Percival et al. (2000)).

Theorem 2.7 (AR(1) case)**.** *Suppose $\{X_t, t \in \mathbb{Z}\}$ is a causal Gaussian AR(1) time series, i.e.*

$$X_t = a X_{t-1} + e_t, \qquad e_t \overset{i.i.d.}{\sim} N(0, \sigma^2), \qquad |a| < 1,$$

and denote by $\{Y_t\}$ *the corresponding synthesized time series ignoring all dependence between coefficients as given in* (2.15) *and* (2.16) *using a discrete wavelet transform based on the filter* $h_0, \ldots, h_{M-1}$. *Then,*

$$|n\,\mathrm{var}(T_n(\boldsymbol{Y})) - n\,\mathrm{var}(T_n(\boldsymbol{X}))| \not\to 0$$

with the exception of at most finitely many a.

The findings of Theorems 2.5 and 2.7 and Corollary 2.6 are supported by the data analysis in Aston et al. (2005). In this study, the variance of regression parameters (in the wavelet domain) is estimated based both on this independence assumption as well as using exact methods. Their Figure 1 indicates that the approximation assuming independence is unsatisfactory for their fMRI data example.

In the context of parameter estimation for fractionally differenced processes, Craigmile et al. (2005) propose to make use of the discrete wavelet transform to decorrelate the data to be able to apply an approximate maximum likelihood approach. They argue that dependence between levels can be handled by increasing the length M of the wavelet filter, whereas the dependence within a level can be modeled by low-order autoregressive processes. In that spirit, the following subsection addresses whether capturing the dependence within levels while assuming independence between levels has the potential of valid asymptotic results.

2.2. Ignoring only the dependence between coefficients of different levels

In this section, we investigate the influence of the dependence between levels if the dependence of every level is mimicked correctly. More precisely, we consider

$$\begin{aligned}&\widetilde{\boldsymbol{Y}} = \boldsymbol{W}^T \boldsymbol{D}_{\widetilde{\boldsymbol{Y}}}, \\ &\boldsymbol{D}_{\widetilde{\boldsymbol{Y}}} = (d_{\widetilde{\boldsymbol{Y}},1}(1), \ldots, d_{\widetilde{\boldsymbol{Y}},1}(n/2), d_{\widetilde{\boldsymbol{Y}},2}(1), \ldots, d_{\widetilde{\boldsymbol{Y}},2}(n/4), \ldots, d_{\widetilde{\boldsymbol{Y}},D}(1), v_{\widetilde{\boldsymbol{Y}},D}(1))^T,\end{aligned} \tag{2.18}$$

where

$$(d_{\widetilde{\boldsymbol{Y}},k}(1), \ldots, d_{\widetilde{\boldsymbol{Y}},k}(n_k)) \overset{\mathcal{D}}{=} (d_{\boldsymbol{X},k}(1), \ldots, d_{\boldsymbol{X},k}(n_k)), \quad k = 1, \ldots, D.$$

We will concentrate on the consideration of the Gaussian case for dependent data. Nevertheless, it is worth pointing out that for $X_1, \ldots, X_n$ being i.i.d. the corresponding results of Theorem 2.3 remain true for the above synthesized time series for the Haar basis. This is due to the fact that in this case the wavelet coefficients of each level are truly independent by definition; compare (2.14). Consequently, in this special case, the above distribution coincides with the one discussed in the previous section.

Theorem 2.8 (1-dependent case). *Suppose* $\{X_t, t \in \mathbb{Z}\}$ *is Gaussian and 1-dependent with* $\mathrm{E}\, X_t = 0$ *and* $\mathrm{E}\,|X_t|^4 < \infty$ *and denote by* $\{\widetilde{Y}_t\}$ *the corresponding synthesized time series ignoring the dependence between levels as given in* (2.18) *using a discrete wavelet transform based on the filter* $h_0, \ldots, h_{M-1}$.

(a) Then, $\mathrm{E}\, T_n(\widetilde{\boldsymbol{Y}}) = \mathrm{E}\, T_n(\boldsymbol{X})$, *but there exists a constant* $\widetilde{\beta}_{\boldsymbol{W}_n}$ *only depending on* $\boldsymbol{W}_n$ *such that*

$$n\,\mathrm{var}(T_n(\widetilde{\boldsymbol{Y}})) = \left(2 - \frac{2}{n}\right)\gamma_{\boldsymbol{X}}^2(0) + \alpha_n \gamma_{\boldsymbol{X}}(0)\gamma_{\boldsymbol{X}}(1) + \widetilde{\beta}_{\boldsymbol{W}_n}\gamma_{\boldsymbol{X}}^2(1),$$

with $\beta_n - \widetilde{\beta}_{\boldsymbol{W}_n} \geq \frac{1}{4}\left((h_0^{(2)} h_{M-1})^2 + (h_0 h_{3M-3}^{(2)})^2\right) > 0$ *with* $h_0^{(2)}, h_{3M-3}^{(2)}$ *as in* (2.8) *for all* $n \geq 4M$.

(b) If $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$ *is the Haar-wavelet transform, then* $\widetilde{\beta}_{\boldsymbol{W}_{n,H}} \to 20/7$.

The above theorem shows that ignoring only the dependence between levels still leads to asymptotically inconsistent results for any fixed wavelet basis even in the Gaussian case because the asymptotic variance of $T_n(\boldsymbol{X})$ requires a factor 4 in front of $\gamma_{\boldsymbol{X}}^2(1)$ (compare with Theorem 2.1).

In fact, we actually proved in Theorem 2.5 that for $n \geqslant 2M$

$$\beta_n - \beta_{\boldsymbol{W}_n} \geqslant \beta_n - \widetilde{\beta}_{\boldsymbol{W}_n} + (h_0 h_{M-1})^2.$$

Consequently, while taking the dependence within levels into account does not yield consistent results, it does reduce the error. The given lower bounds are very conservative and it is to be expected that the true error is much larger as can e.g. be seen when comparing with the true error of the Haar basis.

Remark 2.2. *For the Haar basis we can now compare the limiting variance in Theorem 2.1(iii) with the corresponding results in Theorem 2.5 and Theorem 2.8. This shows that the factor of* $\gamma_{\boldsymbol{X}}^2(1)$ *indeed improves from* $16/7$ *to* $20/7$ *when the dependence within levels is taken into account. However, there are still* $8/7$ *lacking to the proper factor* 4 *which must consequently be coded in the dependence between levels. In particular, this shows that not only is the dependence between levels not negligible, but its contribution is even twice as large as the one within the levels. Some heuristic explanation for this effect may be that there is dependence not only to one wavelet coefficient of the next finer scale but to two (in the 1-dependent case - otherwise even more). This effect is in contrast to popular believe which is based on the observation that the correlation decreases exponentially between levels. This is because the latter effect is merely an artefact of the factor leading to normalized* $\boldsymbol{h}_{k,j}$.

The main consequence of the above theorem is summarized in the following corollary.

Corollary 2.9. *Let the assumptions of Theorem 2.8 (a) hold and* $\gamma_{\boldsymbol{X}}(1) \neq 0$. *Then,* $\mathrm{E}\, T_n(\widetilde{\boldsymbol{Y}}) = \mathrm{E}\, T_n(\boldsymbol{X})$, *but*

$$d_2\left(\sqrt{n}(T_n(\widetilde{\boldsymbol{Y}}) - \mathrm{E}\, T_n(\widetilde{\boldsymbol{Y}}))\,,\, \sqrt{n}(T_n(\boldsymbol{X}) - \mathrm{E}\, T_n(\boldsymbol{X}))\right) \underset{n\to\infty}{\not\to} 0.$$

Remark 2.3. *The assertion of the above theorem as well as corollary remains true if we synthesize a time series using independent wavelet coefficients of levels 1 to* j ($j \geqslant 2$)*, which are also independent of the remaining scaling coefficients, but mimic the joint distribution of the scaling coefficients correctly. The reason is that the proof is merely based on the two finest levels of wavelet coefficients.*

Similarly to the one-dependent case, the following theorem shows that taking more dependence (between wavelet coefficients within each level) into account still is insufficient also for autoregressive time series of order 1.

Theorem 2.10 (AR(1) case)**.** *Suppose $\{X_t, t \in \mathbb{Z}\}$ is a causal Gaussian AR(1) time series, i.e.*

$$X_t = aX_{t-1} + e_t, \qquad e_t \overset{i.i.d.}{\sim} N(0, \sigma^2), \qquad |a| < 1,$$

and denote by $\{\widetilde{Y}_t\}$ the corresponding synthesized time series ignoring the dependence between levels as given in (2.18) *using a discrete wavelet transform based on the filter $h_0, \ldots, h_{M-1}$. Then,*

$$\left| n \operatorname{var}(T_n(\widetilde{\boldsymbol{Y}})) - n \operatorname{var}(T_n(\boldsymbol{X})) \right| \not\to 0$$

with the exception of at most finitely many a.

Remark 2.4. *While the asymptotic variance of the sample variance for the synthesized time series is asymptotically incorrect due to the lack of dependence between the wavelet coefficients, the precise error depends on the wavelet transform used as well as the time series structure of the data. Depending on the interplay between those two (in particular if data driven methods are used to choose the wavelet transform) the model error (by assuming independence) can become quite small, so that the performance of the procedure in small samples may become better even than say a block bootstrap in the time domain, which has no systematic error but may have a larger small sample error. Since our results are based on the assumption that the same mother wavelet is used for each n, it may even be possible to make the systematic error disappear asymptotically by choosing different mother wavelets for each n adapted to the underlying time series structure in such a way that the systematic error introduced by each of those 'bases' becomes smaller and smaller. This is in the same spirit as the combination of a wavelet transform with decorrelation tests as suggested by Percival et al. (2000) in their wavestrap.*

2.3. Sample autocovariance and sample autocorrelation

In this section, we address the asymptotic behavior of the sample autocovariance and autocorrelation at lag one for synthesized time series $\boldsymbol{Y}$ and $\widetilde{\boldsymbol{Y}}$ as proposed in the previous Sections 2.1 and 2.2. We define

$$\widehat{\gamma}_{\boldsymbol{Y}}(1) = \frac{1}{n} \sum_{t=1}^{n-1} (Y_{t+1} - \bar{Y}_n)(Y_t - \bar{Y}_n), \quad \text{and} \quad \widehat{\rho}_{\boldsymbol{Y}}(1) = \frac{\widehat{\gamma}_{\boldsymbol{Y}}(1)}{\widehat{\gamma}_{\boldsymbol{Y}}(0)},$$

where $\widehat{\gamma}_{\boldsymbol{Y}}(0) := T_n(\boldsymbol{Y})$ and similarly $\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)$ and $\widehat{\rho}_{\widetilde{\boldsymbol{Y}}}(1)$ as well as $\widehat{\gamma}_{\boldsymbol{X}}(1)$ and $\widehat{\rho}_{\widetilde{\boldsymbol{X}}}(1)$ for the original time series $\boldsymbol{X}$. The following theorem proves a systematically erratic bias for $\widehat{\gamma}_{\boldsymbol{Y}}(1)$ and $\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)$.

Theorem 2.11 ($\widehat{\gamma}_{\boldsymbol{Y}}(1)$ and $\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)$, 1-dependent case)**.** *Suppose $\{X_t, t \in \mathbb{Z}\}$ is Gaussian and 1-dependent with $\operatorname{E} X_t = 0$, $\operatorname{E}|X_t|^2 < \infty$ and $\gamma_{\boldsymbol{X}}(1) \neq 0$. Denote by $\{Y_t\}$ and $\{\widetilde{Y}_t\}$ the corresponding synthesized time series as given in* (2.15) *and* (2.18)*, respectively, using a discrete wavelet transform based on the filter $h_0, \ldots, h_{M-1}$. β_n is defined in* (2.17)*.*

(a) With $\{Y_t\}$ and $\beta_{\boldsymbol{W}_n}$ as in Theorem 2.5, we have

(i) $|\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1)| = \frac{1}{4}(\beta_n - \beta_{\boldsymbol{W}_n})|\gamma_{\boldsymbol{X}}(1)| \geq \frac{1}{4}(h_0 h_{M-1})^2 |\gamma_{\boldsymbol{X}}(1)| > 0$ *for all* $n \geq 2M$.

(ii) *If* $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$, *we have* $\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1) \to \frac{4}{7}\gamma_{\boldsymbol{X}}(1)$.

(b) *With* $\{\widetilde{Y}_t\}$ *and* $\widetilde{\beta}_{\boldsymbol{W}_n}$ *as in Theorem 2.8, we have*

(i) $|\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)| = \frac{1}{4}(\beta_n - \widetilde{\beta}_{\boldsymbol{W}_n})|\gamma_{\boldsymbol{X}}(1)| \geq \frac{1}{4}\left((h_0^{(2)} h_{M-1})^2 + (h_0 h_{3M-3}^{(2)})^2\right)|\gamma_{\boldsymbol{X}}(1)| > 0$ *with* $h_0^{(2)}, h_{3M-3}^{(2)}$ *as in* (2.8) *for all* $n \geq 4M$.

(ii) *If* $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$, *we have* $\mathrm{E}\,\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1) \to \frac{5}{7}\gamma_{\boldsymbol{X}}(1)$.

The subsequent corollary gives a similar result for $\widehat{\rho}_{\boldsymbol{Y}}(1)$.

Corollary 2.12 ($\widehat{\rho}_{\boldsymbol{Y}}(1)$ and $\widehat{\rho}_{\widetilde{\boldsymbol{Y}}}(1)$, 1-dependent case)**.** *Under the corresponding assumptions of Theorem 2.11, we have*

(a) (i) $|\rho_{\boldsymbol{X}}(1) - \widehat{\rho}_{\boldsymbol{Y}}(1)| = \frac{1}{4}(\beta_n - \beta_{\boldsymbol{W}_n})|\gamma_{\boldsymbol{X}}(1)|/\gamma_{\boldsymbol{X}}(0) + o_P(1)$.

(ii) *If* $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$, *we have* $\widehat{\rho}_{\boldsymbol{Y}}(1) \xrightarrow{P} \frac{4}{7}\rho_{\boldsymbol{X}}(1)$.

(b) (i) $|\rho_{\boldsymbol{X}}(1) - \widehat{\rho}_{\widetilde{\boldsymbol{Y}}}(1)| = \frac{1}{4}(\beta_n - \widetilde{\beta}_{\boldsymbol{W}_n})|\gamma_{\boldsymbol{X}}(1)|/\gamma_{\boldsymbol{X}}(0) + o_P(1)$.

(ii) *If* $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$, *we have* $\widehat{\rho}_{\boldsymbol{Y}}(1) \xrightarrow{P} \frac{5}{7}\rho_{\boldsymbol{X}}(1)$.

The results of Theorem 2.11 and Corollary 2.12 indicate that the synthesized time series $\boldsymbol{Y}$ and $\widetilde{\boldsymbol{Y}}$ do not carry enough (serial) dependence structure to mimic correctly the means of sample autocovariances and autocorrelations. This is in contrast to the sample variance, where its mean remains correct, but its variance is distorted.

2.4. Relation to classical time series results

The above results have several connections to the following three well-known results from classical time series analysis:

2.4.1. Frequency domain decorrelation and bootstrap

First, wavelet analysis has strong connections to frequency domain inference, for which the discrete Fourier coefficients are (in some sense) asymptotically independent and normal. For a synthesized time series based on independent (and in the i.i.d. case identically distributed) Fourier coefficients the same restrictions apply as in Theorem 2.3 or Corollary 2.4, because the term including $\eta-3$ of the variance vanishes asymptotically for the bootstrap (cf. Dahlhaus and Janas (1996)). This issue has at least been partially solved by the use of hybrid (i.e. time and frequency domain) bootstrap procedures (cf. Kreiss and Paparoditis (2003), Jentsch and Kreiss (2010)). However, frequency domain bootstrap methods have been successfully used in many applications such as the estimation of spectral densities, spectral ratios (cf. Paparoditis (2002) for a review) or even time domain statistics such as change-point or unit-root tests (cf. Kirch and Politis (2011)). Bootstrap consistency is achieved for Gaussian data (in which case $\eta = 3$ and the corresponding term vanishes) as well as in the non-Gaussian case for statistics whose asymptotic distributions depend only on the first and second order structure of the time series.

In case of the frequency domain bootstrap the problem is that some of the dependency information of the original time series is coded in the dependency between Fourier ordinates more precisely in the fourth-order structure. As a result a bootstrap (or equivalently synthesis) that assumes independence between Fourier coefficients only captures the first and second-order structure of the time series correctly (as well as that part of the fourth-order structure that is already explained by the second-order structure). This explains why the method works quite generally for Gaussian data. For non-Gaussian data, the covariance between periodogram ordinates (i.e. the fourth order moment structure of the original process) disappears asymptotically but only with rate $1/n$, which means that in some situations (such as the sample variance) those errors add up (as a sum of n summands of order $1/n$ is no longer asymptotically negligible) while in others (such as the autocorrelation) this effect does not cause a problem so that no Gaussianity assumption needs to be made for the sample autocorrelation and frequency domain methods.

In case of the wavelet bootstraps similar things happen in the non-Gaussian case as pointed out by the factor λ_{O_n} in Theorem 2.3 for i.i.d. data, which is smaller than 1 leading to an analogous systematic error as for frequency domain bootstrap methods and the sample variance. In fact, as λ_{O_n} can be close to 1 if bounded away from it (depending on the choice of wavelet basis and the underlying time series structure), this is an improvement over frequency domain bootstrap methods, where in the above notation $\lambda_{O_n} = 0$. This is no problem in the Gaussian case, where $\eta - 3 = 0$ and it may be true that the term will also disappear in the wavelet domain for the sample autocorrelations for linear processes (analogously to the frequency domain).

However, unlike in the frequency domain, even under Gaussianity assumption we get an **additional** systematic error in the asymptotic variance of the sample variance that relates to the second-order rather than the fourth-order dependence structure (see Theorems 2.5 – 2.10). This error is due to the missing dependence between the wavelet coefficient showing that this dependence is not asymptotically negligible in general. This is fundamentally different from frequency domain bootstrapping.

While frequency domain methods are consistent for the sample autocorrelation for linear processes even in the non-Gaussian case, wavelet based methods introduce a systematic bias even for Gaussian data if dependence between or within the scales is ignored (see Section 2.3). The same is true for the sample autocovariance where also a systematic error is introduced. This is surprising and very different from frequency domain methods as the error for the latter always relates to the asymptotic variance only.

2.4.2. Time domain decorrelation and bootstrap

The famous Wold theorem states that all reasonable stochastic time series can be written as a weak linear process with uncorrelated noise. On the other hand for many statistical results (such as the linear process bootstrap; cf. Jentsch and Politis (2013) and Jentsch and Politis (2015)) the assumption of independence (i.e. strict white noise) is needed, which is far more restrictive. However, for Gaussian data as well as several statistics of interest the asymptotics of the weak and strict linear process coincide.

Thirdly, many processes can be well approximated by an AR(∞) process (cf. Theorem 4.4.3 in Brockwell and Davis (1991)). This approximation has been exploited by the AR-sieve-bootstrap, where Kreiss et al. (2011) investigate bootstrap consistency for a large class of statistics for which the bootstrap turns out to be consistent if and only if the asymptotic distribution of the corresponding statistics only depends on the first and second order structure of the underlying time series. In particular, this implies bootstrap consistency in the Gaussian case.

In view of these classic results, it is very surprising that the approximation via independent wavelet coefficients (or only between-scale independence) does not even yield consistent results in the Gaussian case.

2.4.3. Locally stationary wavelet models

Nason et al. (2000) introduce and investigate locally stationary wavelet models, where a nondecimated wavelet transform is used and the corresponding coefficients are assumed to be uncorrelated. Their Proposition 3 proves that all SRD time series are contained in this model class. Some statistical theory for this class is provided in Fryzlewicz and Nason (2006). Applications range from forecasting in Fryzlewicz et al. (2003), via classification in Fryzlewicz and Ombao (2009) to change point detection in Cho and Fryzlewicz (2012, 2014), Killick et al. (2013) and Nam et al. (2014). For most of the theoretic statistical analysis, however, the stronger assumption of independence (and often even Gaussianity) of the increment process is necessary. In light of the above discussion it may be well worth investigating how strong this assumption truly is and what time domain representations of time series are consistent with it.

3. Bootstrap in the wavelet domain

In this section, we focus on the sample variance and look at the validity of nonparametric bootstrap methods for the wavelet coefficients if more and more dependence is taken into account. In contrast to most of the results of the previous section, the results are not restricted to the 1-dependent case and allow in particular for LRD time series.

3.1. Levelwise i.i.d. bootstrap

In this subsection, we investigate the validity of Efron's standard i.i.d. bootstrap applied independently to each level, i.e. $\{d_k^*(l) : k, l\}$ are i.i.d. with $d_k^*(l)$ uniformly distributed on the set $\{d_k(1), \ldots, d_k(n_k)\}$ (independent of $\{X_t\}$). This is equivalent to $d_k^*(l) \sim \widehat{F}_k$ with $\widehat{F}_k(x) := \frac{1}{n_k}\sum_{l=1}^{n_k} 1_{\{d_{\boldsymbol{X},k}(l)\leqslant x\}}$.

$$\begin{aligned} \boldsymbol{X}^*_{\boldsymbol{W}} &= \boldsymbol{W}^T \boldsymbol{D}^*_{\boldsymbol{X}}, \qquad (3.1) \\ \boldsymbol{D}^*_{\boldsymbol{X}} &= (d_1^*(1), \ldots, d_1^*(n_1), d_2^*(1), \ldots, d_2^*(n_2), \ldots, d_D^*(1), v_{\boldsymbol{X},D(1)})^T. \end{aligned}$$

For the statistic $T_n(\boldsymbol{X})$ the choice of the scaling coefficient is arbitrary. However, for other statistics it may be important that it correctly mimics the asymptotic behavior of the sample mean, which does require additional resampling such as moving blocks resampling in the time domain or subsampling.

This bootstrap method asymptotically mimics the distribution of $\boldsymbol{Y} = \boldsymbol{W}^T \boldsymbol{D}_{\boldsymbol{Y}}$ with $\boldsymbol{D}_{\boldsymbol{Y}}$ as in (2.16), which is related to the concept of a companion process as defined in Kreiss et al. (2011). Consequently, we get the same restriction that we get for the synthesized time series $\boldsymbol{Y}$.

Theorem 3.1. *Define* $\boldsymbol{X}^*_{\boldsymbol{W}}$ *as in* (3.1) *and* $\boldsymbol{Y}$ *as in* (2.15) *and* (2.16).

(a) Let $\{X_t, t \in \mathbb{Z}\}$ *be i.i.d. Then, it holds* $\mathrm{E}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}})) = T_n(\boldsymbol{X})$ *and*

$$\mathrm{E}\left(n \ \mathrm{var}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}}))\right) = n \ \mathrm{var}(T_n(\boldsymbol{Y})) + o(1)$$

(b) Let $\{X_t, t \in \mathbb{Z}\}$ *be Gaussian and* $\gamma_{\boldsymbol{X}}(r) = O(|r|^{2d-1})$ *as* $|r| \to \infty$ *with* $0 \leqslant d < 1/4$. *Then, it holds* $\mathrm{E}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}})) = T_n(\boldsymbol{X})$ *and*

$$\mathrm{E}\left(n \ \mathrm{var}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}}))\right) = n \ \mathrm{var}(T_n(\boldsymbol{Y})) + o(1) \quad \text{if } 0 \leqslant d < 1/4$$

as well as

$$\mathrm{E}\left(\frac{n}{\log n} \ \mathrm{var}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}}))\right) = \frac{n}{\log n} \ \mathrm{var}(T_n(\boldsymbol{Y})) + o(1) \quad \text{if } d = 1/4.$$

The first part of the above theorem in combination with Theorem 2.3 shows that even for i.i.d. data such a bootstrap will fail except for $\eta = 3$ (as for Gaussian data). An analogous assertion also holds if one applies Efron's i.i.d. bootstrap after any orthonormal transformation $\boldsymbol{O}_n$ that is not a permutation matrix. In the latter case, the bootstrap is simply equivalent to Efron's i.i.d. bootstrap of the original data and as such asymptotically valid for i.i.d. data. While assertion (a) is purely academic, after all why would one want to apply such a complicated bootstrap for a simple i.i.d. data set, we see in (b) that this bootstrap will not be valid in general for the dependent Gaussian case, because Theorem 2.5 shows that the missing dependency between coefficients is not asymptotically negligible in general. For $d > 1/4$ the proof of Theorem 3.1 does not work and, moreover, Theorem 4.1 below shows that the above bootstrap does not yield consistent results in this case. Since the asymptotic variance of $\boldsymbol{Y}$ differs in general from that of $\boldsymbol{X}$ by Theorem 2.3, the bootstrap variance will typically differ from the true variance of our statistic. These findings are supported by several simulation studies conducted in Percival et al. (2000), Breakspear et al. (2003) and in particular by Tang et al. (2008), who find wavestrap procedures based on nonparametric resampling methods unreliable.

3.2. Levelwise block bootstrap

A non-parametric bootstrap alternative which takes some of the dependency into account is a levelwise (non-overlapping) block bootstrap. To investigate this approach, define the block length for level k by $L_k = 2^{b_k} \leqslant n_k = n/2^k$ for some suitable integer-valued, non-negative $b_k = b_k(n)$ and draw $N_k = n_k/L_k = 2^{D-k-b_k} \geqslant 1$ independent blocks of length L_k uniformly (with replacement) from

$$\{d_k(1),\ldots,d_k(L_k)\}, \quad \{d_k(L_k+1),\ldots,d_k(2L_k)\}, \quad \ldots \quad ,\{d_k\left(n_k-L_k+1\right),\ldots,d_k\left(n_k\right)\}.$$

The kth level of bootstrap coefficients is then given by

$$\widetilde{d}_k^*(1),\ldots,\widetilde{d}_k^*(L_k), \quad \widetilde{d}_k^*(L_k+1),\ldots,\widetilde{d}_k^*(2L_k), \quad \ldots \quad ,\widetilde{d}_k^*\left(n_k-L_k+1\right),\ldots,\widetilde{d}_k^*\left(n_k\right).$$

This is done independently for each level k and independent of $\{X_t\}$ leading to the bootstrap time series

$$\begin{aligned} &\widetilde{\boldsymbol{X}}_{\boldsymbol{W}}^* = \boldsymbol{W}^T \widetilde{\boldsymbol{D}}_{\boldsymbol{X}}^*, \\ &\widetilde{\boldsymbol{D}}_{\boldsymbol{X}}^* = (\widetilde{d}_1^*(1),\ldots,\widetilde{d}_1^*(n_1),\widetilde{d}_2^*(1),\ldots,\widetilde{d}_2^*(n_2),\ldots,\widetilde{d}_D^*(1),v_{\boldsymbol{X},D}(1))^T. \end{aligned} \tag{3.2}$$

This bootstrap method asymptotically mimics the distribution of $\widetilde{\boldsymbol{Y}} = \boldsymbol{W}^T\widetilde{\boldsymbol{D}}_{\boldsymbol{Y}}$ as in (2.18). In order to simplify the conditions on the levelwise bandwidth b_k we assume the existence of a sequence $A(n) \to \infty$ such that as $n \to \infty$

$$\inf_{k \leqslant A(n)} b_k \to \infty, \qquad \sup_{k \leqslant A(n)} 2^{b_k+k}/n \to 0. \tag{3.3}$$

The first condition says that the bandwidth converges to infinity in some uniform sense, where one cannot expect it to hold uniformly over all levels because the number of coefficients $n_k = n/2^k$ belonging to the coarser scales (k close to D) does not increase asymptotically. The second condition corresponds to the second standard condition of block bootstrapping: $L_k/n_k \to 0$.

A possible choice that is (in some sense) uniform in all levels is given by $b_k = \max(b-k+1,0)$ with $b = b(n) \to \infty$ and $2^b/n \to 0$. This choice will play an important role in the next section. In particular, it fulfills the above assumptions (3.3) with $A(n) = \sqrt{b}$.

While this improves the approximation in comparison to the full i.i.d. bootstrap of the previous section, the bootstrap is not asymptotically valid in general due to too much dependence between levels that is not taken into account. In fact, instead of correctly mimicking the asymptotic variance of the process $\{X_t\}$ the bootstrap mimics the process $\{\widetilde{Y}_t\}$.

Theorem 3.2. *Let $\{X_t, t \in \mathbb{Z}\}$ be Gaussian and $\gamma_{\boldsymbol{X}}(r) = O(|r|^{2d-1})$ as $|r| \to \infty$ with $0 \leqslant d < 1/4$ and define $\widetilde{\boldsymbol{X}}_{\boldsymbol{W}}^*$ as in (3.2) and $\widetilde{\boldsymbol{Y}}$ as in (2.18). For block-lengths fulfilling (3.3), it holds $\mathrm{E}^*(T_n(\boldsymbol{X}_{\boldsymbol{W}}^*)) = T_n(\boldsymbol{X})$ and*

$$\mathrm{E}\left(n \ \mathrm{var}^*(T_n(\boldsymbol{X}_{\boldsymbol{W}}^*))\right) = n \ \mathrm{var}\left(T_n(\widetilde{\boldsymbol{Y}})\right) + o(1).$$

For $d = 1/4$ a similar assertion as in Theorem 3.1 can be shown under stronger assumptions on the bandwidth, but for $d > 1/4$ the below proof does not work and Theorem 4.1 shows that the above bootstrap does not yield consistent results in this case.

Remark 3.1. The proof of the theorem remains true for non-Gaussian time series as long as the rates for the covariances of the wavelet coefficients as in Lemma B.1 for $k_1 = k_2$ remain valid. This is in particular the case for all m-dependent time series (m arbitrary but fixed).

Despite the fact that the correlation within levels is now captured correctly by the bootstrap, Theorem 3.2 shows that the asymptotic bootstrap variance will in general differ from the true variance of the statistic. These theoretical findings are supported in parts by the simulations conducted in Sabatini (1999).

3.3. Quadrature Block Bootstrap

The bootstrap procedures in the above two sections fail because they are not able to fully capture the dependence between wavelet coefficients. The question arises whether it is possible to define a block bootstrap which can capture this dependence, where the main problem is the triangular lattice structure of the wavelet coefficients. Additionally, we would like the bootstrap coefficients in each level to be obtained from the set of coefficients of that same level rather than to exchange coefficients between levels. This is particularly important with a possible application to LRD data in mind. In this case, the coefficients are heteroscedastic (between levels) where for an appropriate rescaling the long-memory parameter needs to be known or estimated, something that Percival et al. (2000) explicitly try to avoid by proposing wavestrapping.

To this end, we propose the following bootstrap: First, extend the triangular lattice $\{d_{\boldsymbol{X},k}(l), k = 1,\dots,D, l = 1,\dots,n/2^k\}$ to a rectangular lattice $Q = \{q_{\boldsymbol{X},k}(l) : k = 1,\dots,D, l = 1,\dots,n/2\}$ by duplicating the observed values as shown in Figure 3.1: For level $k = 2$ we write down every coefficient twice, for level $k = 3$ we write it down four times until finally for level $k = D$ the single coefficient is repeated $n/2$ times, that is, we write down the coefficients 2^{k-1} times for level k. Mathematically, this means

$$q_{\boldsymbol{X},k}(l) = d_{\boldsymbol{X},k}\left(\left\lceil \frac{l}{2^{k-1}} \right\rceil\right),$$

where $\lceil x \rceil$ denotes the smallest integer greater or equal to x. The triangular lattice can be obtained from this rectangular lattice by the relation $d_{\boldsymbol{X},k}(l) = q_{\boldsymbol{X},k}(2^{k-1}l)$.

The rectangular lattice can now be block bootstrapped in the following way: Cut Q into vertical (through all levels $k = 1,\dots,D$) stripes of horizontal block length $L = 2^b$ for some suitable integer-valued, non-negative $b = b(n)$. Thus, we obtain $N = \frac{n}{2L}$ such blocks. Now, draw randomly from these blocks (with replacement, independent of $\{X_t\}$) and put them back together to get $q_k^*(l)$, $k = 1,\dots,D$, $l = 1,\dots,n/2$. The bootstrap wavelet coefficients are obtained as $\check{d}_k^*(l) = q_k^*(2^{k-1}l)$. We consider this non-overlapping block bootstrap for simplicity only. Alternatively, an overlapping version can be used which is theoretically more difficult to analyze.

Observe that the effective block length in each level is given by $\max(2^{b-k+1}, 1)$. In particular, as soon as $k \geqslant b + 1$, i.e. for the coarsest scales, this reduces to Efron's i.i.d. bootstrap as described in Section 3.1. However, the impact of those coarsest scales is negligible in the SRD case or for long-memory time series with $d < 1/4$, but will cause

			$d_4(1)$				
	$d_3(1)$				$d_3(2)$		
$d_2(1)$		$d_2(2)$		$d_2(3)$		$d_2(4)$	
$d_1(1)$	$d_1(2)$	$d_1(3)$	$d_1(4)$	$d_1(5)$	$d_1(6)$	$d_1(7)$	$d_1(8)$

$d_4(1)$	$d_4(1)$	$d_4(1)$	$d_4(1)$	$d_4(1)$	$d_4(1)$	$d_4(1)$	$d_4(1)$
$d_3(1)$	$d_3(1)$	$d_3(1)$	$d_3(1)$	$d_3(2)$	$d_3(2)$	$d_3(2)$	$d_3(2)$
$d_2(1)$	$d_2(1)$	$d_2(2)$	$d_2(2)$	$d_2(3)$	$d_2(3)$	$d_2(4)$	$d_2(4)$
$d_1(1)$	$d_1(2)$	$d_1(3)$	$d_1(4)$	$d_1(5)$	$d_1(6)$	$d_1(7)$	$d_1(8)$

Figure 3.1: Schematic representation of the transformation from a triangular (left) to a rectangular (right) lattice for $n = 16$

problems for LRD time series with $d > 1/4$. This is discussed in more detail in Section 4 below.

The bootstrap time series is given by

$$\check{\boldsymbol{X}}^*_{\boldsymbol{W}} = \boldsymbol{W}^T \check{\boldsymbol{D}}^*_{\boldsymbol{X}}, \tag{3.4}$$
$$\check{\boldsymbol{D}}^*_{\boldsymbol{X}} = (\check{d}^*_1(1), \ldots, \check{d}^*_1(n/2), \check{d}^*_2(1), \ldots, \check{d}^*_2(n/4), \ldots, \check{d}^*_D(1), v_{\boldsymbol{X},D}(1))^T.$$

This is essentially the same idea as has been proposed by Wendt et al. (2009) for bootstrapping wavelet leaders of 2D images.

The quadratic scheme above is only an artificial construction to solve the problem of constructing a two-dimensional block bootstrap for the triangular wavelet scheme and should not be confused with the quadratic scheme arising from the use of non-decimated wavelet transforms. For the latter the independence assumption between wavelet coefficients is not even approximately fulfilled for i.i.d. data in the time domain so that the full dependency needs to be captured by any bootstrap method. While a 2-dimensional block bootstrap can be applied quite naturally in this situation, it is unclear how to backtransform the data to the time domain in order to get an approximation of the statistic of interest (or equivalently, rewrite the time domain statistic in terms of the non-decimated wavelet coefficients).

The following theorem shows, that the above bootstrap is in fact capable of correctly capturing the asymptotic variance of SRD as well as LRD (with $d < 1/4$) Gaussian time series of the sample variance in expectation. We believe that these additional assumptions on the basis can be relaxed under stronger assumptions on the rate of decay for the autocovariance function. For example no restrictions apply in the MA(1) case.

Furthermore we conjecture that asymptotic normality could be proved implying asymptotic validity of this bootstrap scheme for a large class of time series. However, a standard block bootstrap in the time domain also yields valid results with good small sample properties for $d < 1/4$, so that this complicated bootstrap does seem to be a bit of an overkill. The next section will show that neither this bootstrap nor most other bootstrap methods will be able to correctly mimic the asymptotic behavior for $d > 1/4$.

Theorem 3.3. *Let $\{X_t, t \in \mathbb{Z}\}$ be Gaussian and define $\check{\boldsymbol{X}}^*_{\boldsymbol{W}}$ as in (3.4). Let $\gamma_{\boldsymbol{X}}(r) = O(|r|^{2d-1})$, $0 \leqslant d < 1/4$, as $|r| \to \infty$ and the basis fulfills for all $k_1 \leqslant k_2$ and all $r \in \mathbb{Z}$ that*

$$\sum_{s=0}^{M_{k_1}-1} (h^{(k_2)}_{s-r})^2 = O(2^{k_1-k_2}), \tag{3.5}$$

which is in particular fulfilled for the Haar basis. Then, $\mathrm{E}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}})) = T_n(\boldsymbol{X})$ *and*

$$\mathrm{E}\left(n\ \mathrm{var}^*(T_n(\check{\boldsymbol{X}}^*_{\boldsymbol{W}}))\right) = n\ \mathrm{var}\,(T_n(\boldsymbol{X})) + o(1),$$

if $L \log n/n \to 0$ *as well as* $L^{4d-2}\, n \log n \to 0$.

Remark 3.2. The assertion remains true for a non-Gaussian time series as long as the rates of decay for the covariances of the wavelet coefficients as proven in Lemma B.1 for the Gaussian case remain true. This is in particular correct for the m-dependent case (m arbitrary but fixed). In particular, this shows that this complicated bootstrap scheme is able to correctly mimic the asymptotic behavior for an i.i.d. sequence with $\eta \neq 3$, which is neither possible via the i.i.d. wavelet bootstrap from Section 2.1 even if the time series is i.i.d. nor via standard frequency domain bootstrap methods (but can easily be achieved by Efrons bootstrap in the time domain for i.i.d. data).

3.4. Simulations

In this section, we aim to illustrate the theoretical properties of synthesis and bootstrap procedures discussed in this paper. More precisely, we apply the DWT-Bootstrap variants of a levelwise iid bootstrap (DWTiid), a levelwise block bootstrap (DWTblock) and a quadrature block bootstrap (DWTquad) as proposed in Sections 3.1 - 3.3 and a time-domain moving block bootstrap (MBB) to simulated data. We consider the statistics of the sample variance, sample autocovariance at lag 1 and the sample autocorrelation at lag 1 based on two short-range dependent (SRD) and two long-range dependent (LRD) time series models

(I) MA(1) $X_t = 0.9e_{t-1} + e_t$,

(II) AR(1) $X_t = 0.9X_{t-1} + e_t$,

(III) FARIMA$(1, d = 0.2, 0)$ with $a = 0.9$,

(IV) FARIMA$(1, d = 0.45, 0)$ with $a = 0.9$,

where we have used standard Gaussian white noise $e_t \sim (0, 1)$ in all cases. We choose SRD Gaussian MA(1) and AR(1) models to illustrate our findings concerning erroneous variances and means for sample autocovariances and autocorrelations in Section 2 for the synthesis and in Sections 3.1 - 3.3 for the corresponding bootstraps. The two LRD FARIMA models cover the cases of an LRD process with moderate ($0 < d < 1/4$) long memory and with strong ($1/4 < d < 1/2$) long memory. This refers to the results established in Section 4 below, where the general inability of DWT-based bootstrap for time series with a strong long range dependence is discussed.

In general, we aim to illustrate distortions for DWTiid and DWTblock for fixed wavelet bases, but also we want to indicate the potential of DWTquad to lead to correct approximations. To compute the DWT for all bootstraps, we used the **R** command *wd(x,...)* with (optional) parameters $family.number \in \{1, 2, 4\}$ (e.g. 1 refers to Haar wavelets) and $family = "DaubExPhase"$, in order to study the effect of different wavelet bases used for the DWT to decorrelate the data. To check whether a higher degree of decorrelation might lead to valid results and whether this might outperform well-established time-domain approaches, we compare our results to MBB, compare also the discussion in Remark 2.4.

For estimating the sample variance in Figure 3.2 and the sample mean in Figure 3.3 in all bootstrap procedures, we have generated $M = 500$ time series of models (I) - (IV) of sample size $n = 2^{12}$ and $B = 500$ bootstrap replications have been used in each step. To estimate the true target values that are indicated by red lines in Figures 3.2 and 3.3, we have used 20 000 simulations. For the time-domain MBB, we have used the **R** routine *b.star()* to estimate the block length. For the wavelet-domain level-wise block bootstrap, we have used the block lengths 2^{b+1-k} with $b = round(D^{3/4})$ for level k and for the quadrature block-bootstrap, we have used the block length 2^b with $b = round(D^{3/4})$. This choice matches the requirements of Section 3.

Figure 3.2 clearly indicates the distortion for the bootstrap variance estimates based on DWTiid and DWTblock. In contrast, DWTquad tends to be clearly superior to the other two wavelet bootstraps for all models under consideration. This is not only true for SRD, but also for LRD time series. However, DWTquad tends to be comparable to the much simpler time-domain MBB with respect to the bias of the estimates only. With respect to the variance of the estimates, MBB clearly outperforms DWTquad. The bad performance of the MBB for sample autocorrelations in the MA model is caused by the use of *b.star()*, which is tailor-made for choosing the block length for the sample mean. In further simulations (not reported in the paper), we found that this rule picks far too small block lengths for sample autocorrelations. A closer inspection of the results for wavelet based bootstrap methods shows that the performance improves if other (obviously more decorrelating) wavelet bases are used. This can be seen particularly for the sample autocorrelations, where the performance for DWTiid and for DWTblock significantly improves. For both LRD time series with moderate and strong long memory, all four bootstrap types are inconsistent for the sample variance and the sample autocovariance. For the sample autocorrelation, the picture becomes somewhat different, but zooming-in shows a distortion here as well.

Supporting the theory in Section 3 the left column of Figure 3.3 shows that the bootstrap means for all bootstrap methods under consideration are not distorted for the statistic of the sample variance. In contrast and as discussed in Section 2.3, the picture is different for the sample autocovariance and sample autocorrelation at lag 1. In particular, DWT-based bootstrap sample autocovariances and autocorrelations tend to systematically underestimate the true ones. Similar to the observations made above in Figure 3.2, this effect tends to vanish by using more suitable wavelet bases and by using more refined DWT bootstraps, i.e. by moving from DWTiid to DWTblock or DWTquad. The time-domain MBB also tends to underestimate the true values particularly for LRD time series, but only slightly (and not systematically) for SRD time series. The latter effect can again be explained by the use of *b.star()* picking a too small block length. The behavior for sample autocovariances and autocorrelations at higher lags is similar in having a systematic downward distortion (the exact results are not reported here).

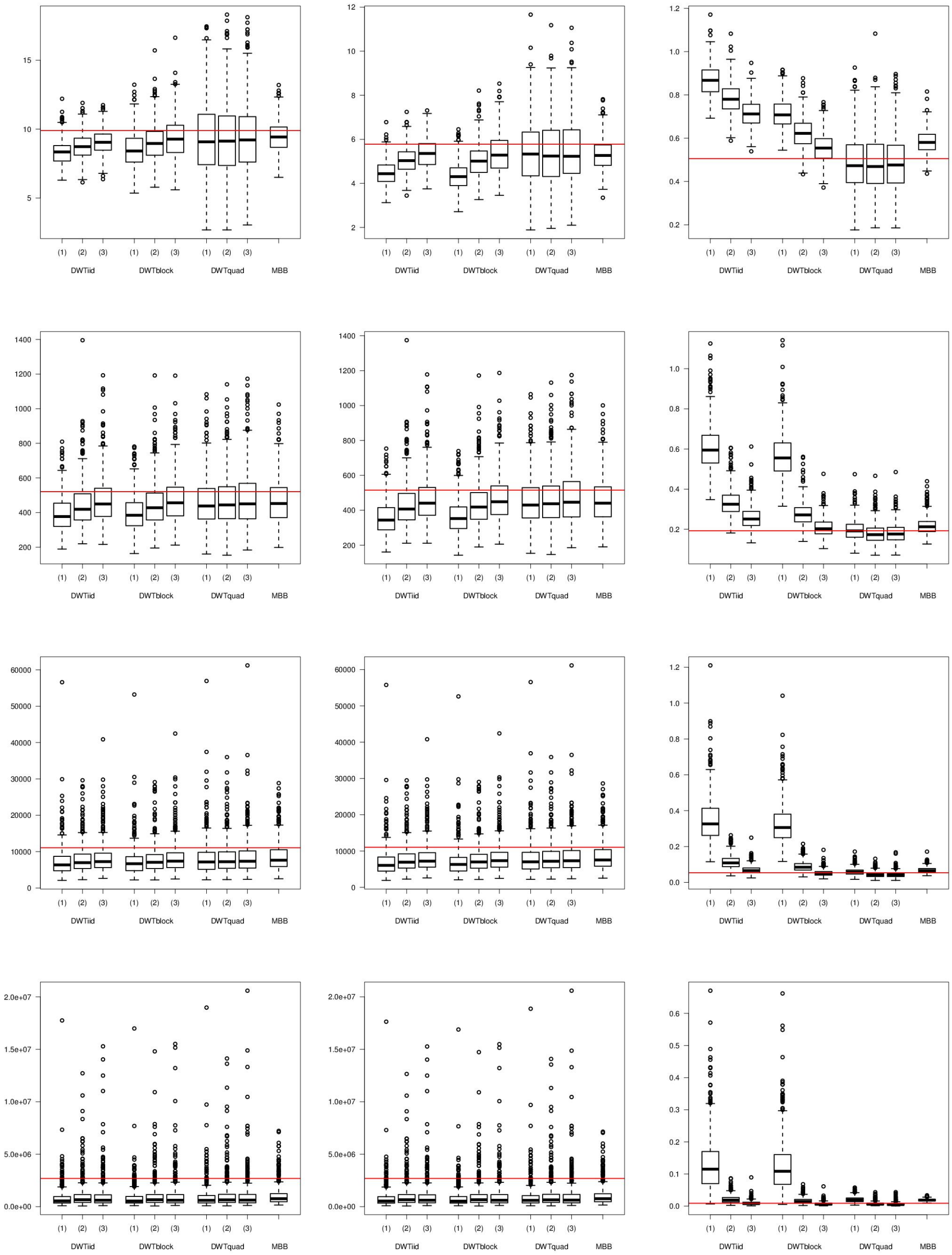

Figure 3.2: Boxplots of bootstrap variance estimates based on DWTiid, DWTblock, DWTquad and MBB for the sample variance (left column), sample autocovariance at lag 1 (center column) and sample autocorrelation at lag 1 (right column). For each DWT-type bootstrap, the three boxes correspond to different wavelet bases with $family.number \in \{1, 2, 4\}$ (from left to right). The results are shown for generated data of sample size $n = 2^{12}$ for models (I) - (IV) (from top to bottom). The targets of the true variances are marked with red horizontal lines.

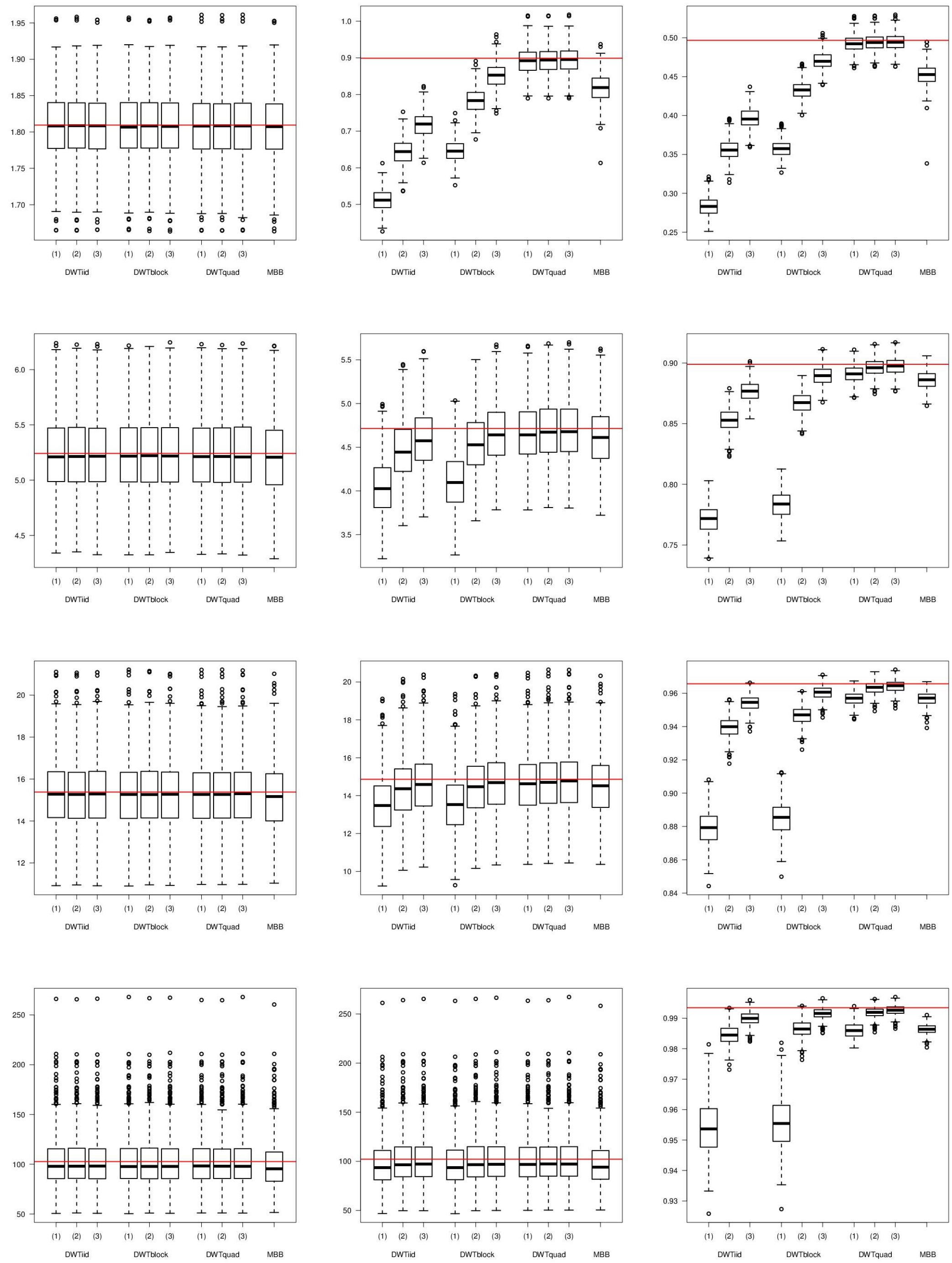

Figure 3.3: Boxplots of bootstrap mean estimates based on DWTiid, DWTblock, DWTquad and MBB for the sample variance (left column), sample autocovariance at lag 1 (center column) and sample autocorrelation at lag 1 (right column). For each DWT-type bootstrap, the three boxes correspond to different wavelet bases with $family.number \in \{1, 2, 4\}$ (from left to right). The results are shown for generated data of sample size $n = 2^{12}$ for models (I) - (IV) (from top to bottom). The targets of the true means are marked with red horizontal lines.

4. Asymptotic impact of the wavelet smooth under long range dependence

The bootstrap procedures considered in the previous section have one thing in common: None of them is capable to correctly mimic the distribution of the wavelet coefficients in the coarsest levels ($k = D, \ldots, D - J$). This is not an artificial property of these particular bootstrap methods but inherent to all bootstrap methods which try to mimic distributional properties of the wavelet coefficients of each level based only on that same information. The reason is that for $k = D, \ldots, D - J$, J fixed, the number of coefficients from which to draw information is only finite (even as $n \to \infty$). It could possibly be solved by taking more information into account, where one possibility is subsampling (in the wavelet domain), which in the context of wavelet coefficients reduces to resampling from a finer level and then use a rescaling step (involving the long-memory parameter d) to put the statistics on the same 'footing'. Using scaling coefficients rather than wavelet coefficients for the coarser scales does not change the problem as can be seen by Theorem 4.1.

In the context of SRD random variables as well as LRD random variables with $d < \frac{1}{4}$ as above, this inability to correctly mimic the coarser scale coefficients did not cause the bootstrap to fail, because the asymptotic influence of those coefficients turns out to be negligible in Theorem 4.1 below.

For LRD data $\{X_t\}$ with long-memory parameter $0 < d < 1/2$, the variance of the kth level coefficients is usually of order 2^{2kd}. For Haar wavelets this can easily be seen by the fact that

$$\operatorname{var}\left(\sum_{t=1}^{m} X_t\right) \sim m \sum_{|h|<m} \gamma_{\boldsymbol{X}}(h) \sim m^{1+2d}$$

and the formula for the Haar wavelet coefficients as in (2.14) and is already stated in (1.9) in Kaplan and Kuo (1993).

The following theorem states that the information in the coarser scale coefficients is only asymptotically negligible if $d \leqslant 1/4$. In fact, for $d > 1/4$ even the single coefficient of the coarsest level (i.e. $k = D$) has an influence on the asymptotic distribution. In this situation the long-range dependence is strong enough to change the asymptotic behavior as can be seen by the different (in comparison to $d \leqslant 1/4$) limit distributions in Theorem 2.2. The influence of the coarse scale information can then be explained by the fact that this information essentially corresponds to the long-range information.

As a consequence, any bootstrap procedure which cannot correctly mimic the distribution of every single wavelet (or scaling) coefficient will fail in general. As already mentioned, one possible way out is to use subsampling for coarser levels. However, to do so we need to know the long-memory parameter d in order to do the proper upscaling. Furthermore, if one does know this parameter (or has a reasonably good estimator) it is much simpler and seems much more efficient to use subsampling on the actual statistic of interest in the time domain; compare Chapter 10 in Lahiri (2003).

Theorem 4.1. *Denote by*

$$T_{n,J}(\boldsymbol{X}) = \frac{1}{n} \sum_{k=1}^{D-J} \sum_{l=1}^{n_k} d_k^2(l)$$

the part of statistic $T_n(\boldsymbol{X})$ *based only on the wavelet rough up to level* $D - J$*. Assume that there exists* $0 < c_d < C_d < \infty$ *such that*

$$c_d\, 2^{2kd} \leqslant \operatorname{var} d_k(\cdot) \leqslant C_d\, 2^{2kd}.$$

Then, it holds for $a(n,d)$ *as in Theorem 2.2:*

a) For $d \leqslant 1/4$*, then*

$$\mathrm{E}\, a(n,d)\, |T_n(\boldsymbol{X}) - T_{n,J}(\boldsymbol{X})| \to 0$$

if $J = J(n) \to \infty$ *sufficiently slow.*

b) For $d > 1/4$*, then*

$$\mathrm{E}\, a(n,d)\, |T_n(\boldsymbol{X}) - T_{n,1}(\boldsymbol{X})| = \mathrm{E}\, \frac{a(n,d)}{n} d_D^2(1) \geqslant c_d > 0.$$

While the quadrature block bootstrap as in Section 3.3 has a chance of being valid as long as $d < 1/4$ with a question mark for $d = 1/4$, analogous results to Theorem 3.3 can be obtained for the much simpler block bootstrap in the time domain for $d < 1/4$. This time domain block bootstrap seems to be preferable to the complicated wavelet based method if only for its simplicity. Kim and Nordman (2011) as well as Lahiri (1993) investigate the validity of the time domain block bootstrap for the sample mean in the presence of LRD.

5. Conclusions

In this paper, we have investigated how much information the dependence between wavelet coefficients carries. It turns out that for the sample variance this dependence is not negligible in general as ignoring it leads to distortions of its limiting variance. Moreover, for sample autocovariances and autocorrelations, we found that already the mean is distorted by ignoring this dependence, which is in contrast to the sample variance, where its mean remains unaffected. Contrary to frequency domain or linear process approximations which yield asymptotically correct results at least for Gaussian time series, this wavelet domain approach turns out to be invalid even for normal data. This is not only true for the dependence between wavelet coefficients of the same level but also for the dependence between wavelet coefficients of different scales. The example of a Haar basis even suggests that the dependence of different levels may contribute more than the dependence within levels. As a consequence, bootstrap methods which are unable to correctly mimic **all** of the dependency structure in the wavelet domain, will be inconsistent. Similarly statistical inference drawn under this assumption will likely contain some model error that is not asymptotically negligible. These theoretical findings are supported by the data analysis in Aston et al. (2005) as well as the simulations in Tang et al. (2008) among others.

While our focus was mainly on the sample variance as a statistic, we indicated also distortions for sample autocovariances and autocorrelations. Consequently, these negative results urge to be extremely careful for other statistics as well. Furthermore, while we only considered the synthesis of time series in addition to bootstrap procedures based on a simplified dependency structure in the wavelet domain, related procedures such as Bayesian inference based on an approximated likelihood (under independence assumptions of the coefficients) should also be handled with care.

Additionally, for LRD time series with long-memory parameter $d > 1/4$, the asymptotic distributional properties of the sample variance depend on the wavelet smooth (or equivalently final scaling coefficients). Because there exist only finitely many of those coefficients (even asymptotically), their distribution cannot be mimicked correctly by standard resampling methods without using additional information about the structure of the time series. Similarly, in the Bayesian analysis the effective sample size for those coefficients is very small, so that the information carried by them will depend mainly on the prior. For these reasons an extra error in the inference for LRD data using those methods cannot be avoided.

Acknowledgements

The research of the first author was supported by the Research Center (SFB) 884 ”Political Economy of Reforms” (Project B6), funded by the German Research Foundation (DFG). Most of this work was done while the second author was still at Karlsruhe Institute of Technology (KIT), where her position was financed by the Stifterverband für die Deutsche Wissenschaft by funds of the Claussen-Simon-trust. We would like to thank Rainer von Sachs during a visit to whom the idea for this work arose. Finally, the paper has greatly benefited from discussions with John Aston, Peter Craigmile, Idris Eckley, Jürgen Franke and Rainer von Sachs.

References

C. Angelini, D. Cava, G. Katul, and B. Vidakovic. Resampling hierarchical processes in the wavelet domain: A case study using atmospheric turbulence. *Physica D*, 207:24–40, 2005.

J.A.D. Aston, R. N. Gunn, R. Hinz, and F. E. Turkheimer. Wavelet variance components in image space for spatiotemporal neuroimaging data. *NeuroImage*, 25(1):159–168, 2005.

J. Beran, Y. Feng, S. Ghosh, and R. Kulik. *Long-Memory Processes - Probabilistic Properties and Statistical Methods*. Springer, 2013.

P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981.

M. Breakspear, M. Brammer, and P.A. Robinson. Construction of multivariate surrogate sets from nonlinear data using the wavelet transform. *Physica D*, 182(1-2):1–22, 2003.

P.J. Brockwell and R.A. Davis. *Time series: theory and methods. 2nd ed.* Springer, 1991.

E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T.A. Carpenter, and M. Brammer. Colored noise and computational inference in neurophysiological (fmri) time series analysis: resampling methods in time and wavelet domains. *Hum Brain Mapp*, 12(2):61–78, 2001.

C. K. Carter and R. Kohn. Semiparametric bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):255–268, 1997.

References

H. Cho and P. Fryzlewicz. Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22:207–229, 2012.

H. Cho and P. Fryzlewicz. Multiple change-point detection for high-dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society Series B*, 2014. To appear.

A. Contreras-Cristán, E. Gutiérrez-Peña, and S.G. Walker. A note on whittle's likelihood. *Communications in Statistics – Simulation and Computation®*, 35(4):857–875, 2006.

P. F. Craigmile and D. B. Percival. Asymptotic decorrelation of between-scale wavelet coefficients. *Information Theory, IEEE Transactions on*, 51(3):1039–1048, 2005.

P. F. Craigmile, P. Guttorp, and D. B. Percival. Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes. *Signal Processing, IEEE Transactions on*, 53(8):3151–3161, 2005.

R. Dahlhaus and D. Janas. A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics*, 24(5):1934–1963, 1996.

I. Daubechies. *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

R. W. Dijkerman and R. R. Mazumdar. On the correlation structure of the wavelet coefficients of frational brownian motion. *IEEE Transactions on Information Theory*, 40(5):1609–1612, 1994.

Y. Fan. On the approximate decorrelation property of the discrete wavelet transform for fractionally differenced processes. *IEEE Transactions on Information Theory*, 49(2):516–521, 2003.

G. Faÿ, E. Moulines, F. Roueff, and M. S. Taqqu. Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics*, 151(2):159–177, 2009.

H. Feng, T. R. Willemain, and N. Shang. Wavelet-based bootstrap for time series analysis. *Communications in Statistics – Simulation and Computation*, 34(2):393–413, 2005.

P Flandrin. Wavelet analysis and synthesis of fractional brownian motion. *IEEE Transactions on Information Theory*, 38(2):910–917, 1992.

P. Fryzlewicz and G. Nason. Haar-fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society, Series B*, 68:611–634, 2006.

P. Fryzlewicz and H. Ombao. Consistent classification of nonstationary time series using nonstationary time series by wavelet process modelling. *Journal of the American Statistical Association*, 104:299–312, 2009.

P. Fryzlewicz, S. Van Bellegem, and R. von Sachs. Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55:737–764, 2003.

J.R.M. Hosking. Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long memory time series. *Journal of Econometrics*, 73:261–284, 1996.

M. J. Jensen. An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control*, 24(3):361–387, 2000.

C. Jentsch and J.-P. Kreiss. The multiple hybrid bootstrap - resampling multivariate linear processes. *Journal of Multivariate Analysis*, 10:2320–2345, 2010.

C. Jentsch and D. N. Politis. Valid resampling of higher order statistics using linear process bootstrap and autoregressive sieve bootstrap. *Communications in Statistics - Theory and Methods*, 42(7):1277–1293, 2013.

C. Jentsch and D. N. Politis. Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *The Annals of Statistics*, 43(3): 1117–1140, 2015.

J. Jeong, M. Vannucci, and K. Ko. A wavelet-based bayesian approach to regression models with long memory errors and its application to fmri data. *Biometrics*, 69(1):184–196, 2013.

L. M. Kaplan and C.-C.J. Kuo. Fractal estimation from noisy data via discrete fractional gaussian noise (dfgn) and the haar basis. *Signal Processing, IEEE Transactions on*, 41(12):3554–3562, 1993.

R. Killick, I. A. Eckley, and P. Jonathan. A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, 7: 1167–1183, 2013.

Y.-M. Kim and D.J. Nordman. Properties of a block bootstrap method under long range dependence. *Sankhya Series A*, 73:79–109, 2011.

C. Kirch and D. N. Politis. TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *The Annals of Statistics*, 39(3):1427–1470, 2011.

K. Ko and M. Vannucci. Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes. *Journal of Statistical Planning and Inference*, 136(10):3415–3434, 2006.

J.-P. Kreiss and S.N. Lahiri. Bootstrap methods for time series. *In: Handbook of Statistics, Vol 30: Time Series-Methods and Applications*, 30:3–26, 2012.

J.-P. Kreiss and E. Paparoditis. Autoregressive-aided periodogram bootstrap for time series. *The Annals of Statistics*, 31(6):1923–1955, 2003.

J.-P. Kreiss and E. Paparoditis. Bootstrap for dependent data: a review. *Journal of the Korean Statistical Society*, 40:357–378, 2011.

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

S. N. Lahiri. On the moving block bootstrap under long range dependence. *Statistics and Probability letters*, 18:405–413, 1993.

S.N. Lahiri. *Resampling methods for dependent data.* Springer, 2003.

J. Liu and P. Moulin. Analysis of interscale and intrascale dependencies between image wavelet coefficients. *in Proc. Int. Conf. Image Processing (ICIP), Vancouver, BC, Canada.*, 2000.

E. J. McCoy and A. T. Walden. Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics*, 5(1):26–56, 1996.

J. Mielniczuk and P. Wojdyllo. Decorrelation of wavelet coefficients for long-range dependent processes. *IEEE Transactions on Information Theory*, 53(5):1879–1883, 2007.

E. Moulines, F. Roueff, and M. S. Taqqu. On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter. *Journal of Time Series Analysis*, 28(2):155–187, 2007a.

E. Moulines, Ff Roueff, and M. S. Taqqu. Central limit theorem for the log-regression wavelet estimation of the memory parameter in the gaussian semi-parametric context. *Fractals*, 15 (04):301–313, 2007b.

E. Moulines, F. Roueff, and M. S. Taqqu. A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series. *The Annals of Statistics*, pages 1925–1956, 2008.

C. F. H. Nam, J. A. D. Aston, I. A. Eckley, and R. Killick. The uncertainty of storm season changes: Quantifying the uncertainty of autocovariance changepoints. *Technometrics*, 2014.

G. Nason. *Wavelet methods in statistics with R.* Springer, 2010.

G. P. Nason, R. von Sachs, and G. Kroisandt. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society, Series B*, 62:271–292, 2000.

E. Paparoditis. Frequency domain bootstrap for time series. *In: H. Dehling, T. Mikosch, M. Sorensen (Eds.), Empirical Process Techniques for Dependent Data, Birkhäuser*, pages 365–381, 2002.

D. B. Percival and D. Mondal. A wavelet variance primer. *In: Handbook of Statistics, Vol 30: Time Series-Methods and Applications*, 30:623–657, 2012.

D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics).* Cambridge University Press, 2000.

D. B. Percival, S. Sardy, and A. C. Davison. Wavestrapping time series: Adaptive wavelet-based bootstrapping. *Nonlinear and Nonstationary Signal Processing*, pages 442–471, 2000.

A. M. Sabatini. Wavelet-based estimation of $1/f$-type signal parameters: Confidence intervals using the bootstrap. *Signal Processing, IEEE Transactions on*, 47(12):3406–3409, 1999.

L. Tang, W. A. Woodward, and W. R. Schucany. Undercoverage of wavelet-based resampling confidence intervals. *Communications in Statistics - Simulation and Computation*, 37(7):1307–1315, 2008.

A. H. Tewfik and M. Kim. Correlation structure of the discrete wavelet coefficients of fractional brownian motion. *IEEE Transactions on Information Theory*, 38(2):904–909, 1992.

B. Vidakovic. *Statistical modeling by wavelets*, volume 503. John Wiley & Sons, 2009.

G. G. Walter. *Wavelets and other orthogonal systems with applications.* CRC Press, 1994.

H. Wendt and P. Abry. Multifractality tests using bootstrapped wavelet leaders. *IEEE Transactions on Signal Processing*, 55(10):4811–4820, 2007.

H. Wendt, P. Abry, and S. Jaffard. Bootstrap for empirical multifractal analysis. *IEEE Signal Processing Magazine*, 24(4):38–48, 2007.

H. Wendt, S. G. Roux, S. Jaffard, and P. Abry. Wavelet leaders and bootstrap for multifractal analysis of images. *Signal Processing*, 89:1100–1114, 2009.

B. Whitcher. Wavelet-based bootstrapping of spatial patterns on a finite lattice. *Computational Statistics & Data Analysis*, 50(9):2399–2421, 2006.

P. Whittle. Curve and periodogram smoothing. *Journal of the American Statistical Association*, 75:122–132, 1957.

G. W. Wornell and A. V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *Signal Processing, IEEE Transactions on*, 40(3):611–623, 1992.

A. Proofs of Section 2

The following lemmas are useful for the below proofs.

Lemma A.1. *Let $T_n(\boldsymbol{X})$ as in (2.1) and suppose the assumptions of Theorem 2.1 are satisfied. Then, for $\boldsymbol{C_X} = \boldsymbol{OX}$, we have*

$$T_n(\boldsymbol{X}) = \frac{1}{n}\boldsymbol{C}_{\boldsymbol{X}}^T\boldsymbol{C_X} - \frac{1}{n^2}\boldsymbol{1}^T\boldsymbol{O}^T\boldsymbol{C_X}\boldsymbol{C}_{\boldsymbol{X}}^T\boldsymbol{O}\,\boldsymbol{1} = \frac{1}{n}\boldsymbol{C}_{\boldsymbol{X}}^T\boldsymbol{C_X} + O_P\left(\frac{1}{n}\right), \tag{A.1}$$

where $\boldsymbol{1} = (1, \dots, 1)^T$. Furthermore, if $\boldsymbol{D_X} = \boldsymbol{WX}$, we have

$$T_n(\boldsymbol{X}) = \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k} d_{\boldsymbol{X},k}^2(l). \tag{A.2}$$

Proof. First, it holds

$$\sum_{t=1}^{n} X_t^2 = \boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{X}^T\boldsymbol{O}^T\boldsymbol{O}\boldsymbol{X} = \boldsymbol{C}_{\boldsymbol{X}}^T\boldsymbol{C_X}, \qquad \bar{X}_n = \frac{1}{n}\boldsymbol{1}^T\boldsymbol{X} = \frac{1}{n}\boldsymbol{1}^T\boldsymbol{O}^T\boldsymbol{C_X}.$$

From this assertion (A.1) follows by $T_n(\boldsymbol{X}) = \frac{1}{n}\sum_{t=1}^{n} X_t^2 - \bar{X}_n^2$ and $\bar{X}_n = O_P(1/\sqrt{n})$. Equation (A.2) follows from

$$\boldsymbol{D}_{\boldsymbol{X}}^T\boldsymbol{D_X} = \sum_{k=1}^{D}\sum_{l=1}^{n_k} d_{\boldsymbol{X},k}^2(l) - v_{\boldsymbol{X},D}^2(1), \qquad \frac{1}{n}v_{\boldsymbol{X},D}^2(1) = \bar{X}_n^2.$$

■

Lemma A.2. *(a) If $\{X_t, t \in \mathbb{Z}\}$ is Gaussian and 1-dependent, it holds*

$$\begin{aligned}&\operatorname{cov}(d_{\boldsymbol{X},k_1}(l_1), d_{\boldsymbol{X},k_2}(l_2))\\ &= \delta_{k_1,k_2}\delta_{l_1,l_2}\gamma_{\boldsymbol{X}}(0) + \left(\sum_{t=1}^{n-1}\left[h_{k_1,l_1}(t)\,h_{k_2,l_2}(t+1) + h_{k_1,l_1}(t+1)\,h_{k_2,l_2}(t)\right]\right)\gamma_{\boldsymbol{X}}(1),\end{aligned}$$

where $\delta_{j,k} = 1$ if $j = k$ and 0 otherwise.

(b) If $\{X_t, t \in \mathbb{Z}\}$ is a causal Gaussian AR(1) time series, i.e.

$$X_t = aX_{t-1} + e_t, \qquad e_t \overset{i.i.d.}{\sim} N(0, \sigma^2), \qquad |a| < 1,$$

it holds

$$\begin{aligned}&\operatorname{cov}(d_{\boldsymbol{X},k_1}(l_1), d_{\boldsymbol{X},k_2}(l_2))\\ &= \gamma_{\boldsymbol{X}}(0)\left[\delta_{k_1,k_2}\delta_{l_1,l_2} + \sum_{j=1}^{n-1} a^j\left(\sum_{t=1}^{n-j}\left[h_{k_1,l_1}(t)\,h_{k_2,l_2}(t+j) + h_{k_1,l_1}(t+j)\,h_{k_2,l_2}(t)\right]\right)\right].\end{aligned}$$

Proof. The proof follows straightforward from

$$d_{\boldsymbol{X},k}(l) = \sum_{t=1}^{n} h_{k,l}(t) X_t$$

and the respective time series structure and is omitted. Note that for an AR(1) time series as in (b) it holds $\gamma(k) = a^k \gamma(0)$. ■

Proof of Theorem 2.3.
Let $\boldsymbol{O}$ be any orthogonal matrix. Similar to (A.1), we have

$$T_n(\boldsymbol{Y}) = \frac{1}{n}\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{C}_{\boldsymbol{Y}} - \frac{1}{n^2}\mathbf{1}^T\boldsymbol{O}^T\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{O}\,\mathbf{1}. \tag{A.3}$$

By construction of $\boldsymbol{C}_{\boldsymbol{Y}}$, it holds $\mathrm{E}\,c_{\boldsymbol{Y}}^2(l) = \mathrm{E}\,c_{\boldsymbol{X}}^2(l) = \gamma_{\boldsymbol{X}}(0)$ and $\mathrm{E}(c_{\boldsymbol{Y}}(l_1)c_{\boldsymbol{Y}}(l_2)) = 0$ for $l_1 \neq l_2$. Because $\{X_t\}$ are i.i.d. and $\boldsymbol{O}$ is orthogonal, $c_{\boldsymbol{X}}(\cdot)$ is uncorrelated, i.e. $\mathrm{E}(c_{\boldsymbol{X}}(l_1)c_{\boldsymbol{X}}(l_2)) = 0$ for $l_1 \neq l_2$. Consequently, $\mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{C}_{\boldsymbol{Y}}\right) = \mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{X}}^T\boldsymbol{C}_{\boldsymbol{X}}\right)$ and $\mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{C}_{\boldsymbol{Y}}^T\right) = \mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{C}_{\boldsymbol{X}}^T\right) = \gamma_{\boldsymbol{X}}(0)\mathbf{Id}$. Together with (A.3), this leads to

$$\begin{aligned}\mathrm{E}\,T_n(\boldsymbol{Y}) &= \frac{1}{n}\mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{C}_{\boldsymbol{Y}}\right) - \frac{1}{n^2}\mathbf{1}^T\boldsymbol{O}^T\,\mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{C}_{\boldsymbol{Y}}^T\right)\boldsymbol{O}\,\mathbf{1}\\ &= \frac{1}{n}\mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{X}}^T\boldsymbol{C}_{\boldsymbol{X}}\right) - \frac{1}{n^2}\mathbf{1}^T\boldsymbol{O}^T\,\mathrm{E}\left(\boldsymbol{C}_{\boldsymbol{X}}\boldsymbol{C}_{\boldsymbol{X}}^T\right)\boldsymbol{O}\,\mathbf{1} = \mathrm{E}\,T_n(\boldsymbol{X}).\end{aligned} \tag{A.4}$$

By the independence of $\{c_{\boldsymbol{Y}}(1), \ldots, c_{\boldsymbol{Y}}(n)\}$ we get $\mathrm{E}(\frac{1}{n^2}\mathbf{1}^T\boldsymbol{O}^T\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{O}\,\mathbf{1})^2 = O(\frac{1}{n^2})$ as well as $\mathrm{E}(\frac{1}{n^2}\mathbf{1}^T\boldsymbol{O}^T\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{O}\,\mathbf{1}) = O(\frac{1}{n})$, which implies

$$\mathrm{var}\left(\frac{1}{n^2}\mathbf{1}^T\boldsymbol{O}^T\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{O}\,\mathbf{1}\right) = O\left(\frac{1}{n^2}\right). \tag{A.5}$$

as well as

$$\frac{1}{n^2}\mathbf{1}^T\boldsymbol{O}^T\boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{O}\,\mathbf{1} = O_P\left(\frac{1}{n}\right). \tag{A.6}$$

As $\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{O}\boldsymbol{X}$, we have $c_{\boldsymbol{X}}(l) = \sum_{p=1}^{n} O(l,p)X_p$ and, consequently, as $\{X_t\}$ is i.i.d. with $\mathrm{E}(X_t^2) = \sigma^2$ and $\mathrm{E}(X_t^4) = \sigma^4\eta$, we have

$$\mathrm{E}(c_{\boldsymbol{X}}^2(l)) = \sum_{p=1}^{n} O^2(l,p)\sigma^2 = \sigma^2, \quad \mathrm{E}(c_{\boldsymbol{X}}^4(l)) = \left(\sum_{p=1}^{n} O^4(l,p)\right)\sigma^4(\eta-3) + 2\sigma^4,$$

which implies

$$\mathrm{var}\,c_{\boldsymbol{Y}}^2(l) = \mathrm{var}\,c_{\boldsymbol{X}}^2(l) = \left(\sum_{p=1}^{n} O^4(l,p)\right)\sigma^4(\eta-3) + 2\sigma^4$$

Furthermore, by (A.5) it holds

$$\begin{aligned}n\,\mathrm{var}(T_n(\boldsymbol{Y})) &= n\,\mathrm{var}\left(\frac{1}{n}\boldsymbol{C}_{\boldsymbol{Y}}^T\boldsymbol{C}_{\boldsymbol{Y}}\right) + O\left(\frac{1}{n}\right) = \frac{1}{n}\sum_{l=1}^{n}\mathrm{var}\;c_{\boldsymbol{Y}}^2(l) + O\left(\frac{1}{n}\right)\\ &= \frac{1}{n}\sum_{l=1}^{n}\sum_{p=1}^{n} O^4(l,p)\,\sigma^4(\eta-3) + 2\sigma^4 + O\left(\frac{1}{n}\right)\\ &=: \lambda_{\boldsymbol{O}_n}\sigma^4(\eta-3) + 2\sigma^4 + O\left(\frac{1}{n}\right).\end{aligned} \tag{A.7}$$

As $\sum_{p=1}^{n} O^2(l,p) = 1$ for all l due to orthogonality, we have $O^2(l,p) \leqslant 1$ for all l and p. Hence, we obtain

$$0 < \lambda_{\boldsymbol{O}_n} = \frac{1}{n} \sum_{l=1}^{n} \sum_{p=1}^{n} O^4(l,p) \leqslant \frac{1}{n} \sum_{l=1}^{n} \sum_{p=1}^{n} O^2(l,p) = 1. \tag{A.8}$$

Furthermore, $\boldsymbol{O}$ is a permutation matrix, iff $\sum_{p=1}^{n} O^4(l,p) = \sum_{p=1}^{n} O^2(l,p) = 1$ holds. This proves assertion (i) of part (a). To show (ii), let $\boldsymbol{O}_n = \boldsymbol{W}_n$ be a discrete wavelet transform based on the filter $h_0, \ldots, h_{M-1}$. Then, similar to (A.2), we get

$$T_n(\boldsymbol{Y}) = \frac{1}{n} \sum_{k=1}^{D} \sum_{l=1}^{n_k} d^2_{\boldsymbol{Y},k}(l), \tag{A.9}$$

leading to

$$\begin{aligned} n\,\mathrm{var}(T_n(\boldsymbol{Y})) &= \left(\frac{1}{n} \sum_{k=1}^{D} \sum_{l=1}^{n_k} \sum_{p=1}^{n} h^4_{k,l}(p) \right) \sigma^4(\eta - 3) + \left(2 - \frac{2}{n}\right) \sigma^4 \\ &=: \lambda_{\boldsymbol{W}_n} \sigma^4(\eta - 3) + \left(2 - \frac{2}{n}\right) \sigma^4. \end{aligned}$$

By splitting $\lambda_{\boldsymbol{W}_n}$ in two parts we get

$$\lambda_{\boldsymbol{W}_n} = \frac{1}{n} \sum_{l=1}^{n_1} \sum_{p=1}^{n} h^4_{1,l}(p) + \frac{1}{n} \sum_{k=2}^{D} \sum_{l=1}^{n_k} \sum_{p=1}^{n} h^4_{k,l}(p). \tag{A.10}$$

The first summand of (A.10) corresponds to $k = 1$, which is the finest level of wavelet coefficients and the second one to the remaining wavelet levels $k = 2, \ldots, D$. As $h^4_{k,l}(p) \leqslant 1$ for all k, l and p, we get $\sum_{p=1}^{n} h^4_{k,l}(p) \leqslant \sum_{p=1}^{n} h^2_{k,l}(p) = 1$ such that the second term in (A.10) can be bounded by $1/2 - 1/n$. For the first summand, we get

$$\frac{1}{n} \sum_{l=1}^{n_1} \sum_{p=1}^{n} h^4_{1,l}(p) = \frac{1}{2} \sum_{l=0}^{M-1} h^4_l < \frac{1}{2}$$

as $h_0, h_{M-1} \neq 0$ and $\sum_{l=0}^{M-1} h_l^2 = 1$ by construction of the DWT in (2.7). This shows (ii). Now, let $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$ be the Haar-wavelet transform. Then, by (2.14), we have 2^{k-1} entries of $2^{-k/2}$ and 2^{k-1} entries of $-2^{-k/2}$ in $\boldsymbol{h}_{k,l}$. This leads to

$$\sum_{p=1}^{n} h^4_{k,l}(p) = 2^{k-1} \left(\left(2^{-k/2}\right)^4 + \left(-2^{-k/2}\right)^4 \right) = 2^{-k},$$

which gives

$$\lambda_{\boldsymbol{W}_{n,H}} = \frac{1}{n} \sum_{k=1}^{D} \sum_{j=1}^{n_k} \sum_{p=1}^{n} h^4_{k,l}(p) = \sum_{k=1}^{D} \frac{1}{2^{2k}} = \frac{1 - (1/4)^{D+1}}{1 - 1/4} - 1 \to \frac{1}{3},$$

proving (iii). Now, we prove the asymptotic normality claimed in part b). Since by (A.6)

$$T_n(\boldsymbol{Y}) = \frac{1}{n}\sum_{l=1}^{n} c^2_{\boldsymbol{Y}}(l) + O_P\left(\frac{1}{n}\right)$$

and as $\{c_{\boldsymbol{Y}}(1), \ldots, c_{\boldsymbol{Y}}(n)\}$ are (a triangular array of) independent random variables by construction, it is sufficient to prove the validity of the Lyapunov-condition. Because X_t^2 is non-degenerate, we have $\operatorname{var}(X_t^2) = \mathrm{E}\, X_t^4 - \left(\mathrm{E}\, X_t^2\right)^2 > 0$, which implies $\eta = \mathrm{E}\, X_t^4/(\mathrm{E}\, X_t^2)^2 > 1$. By (A.7) and (A.8) it holds

$$\frac{1}{n}\sum_{l=1}^{n} \operatorname{var} c^2_{\boldsymbol{Y}}(l) = 2\sigma^4 + \sigma^4(\eta-3)\lambda_{\boldsymbol{O}_n} \geqslant \min(2, \eta-1)\sigma^4 > 0. \tag{A.11}$$

Furthermore, by the independence and centeredness of $\{X_t\}$, we get

$$\begin{aligned}
\mathrm{E}\, c^6_{\boldsymbol{Y}}(l) &= \sum_{p_1,\ldots,p_6=1}^{n} \left(\prod_{j=1}^{6} O(l,p_j)\right) \mathrm{E}\left(\prod_{j=1}^{6} X_{p_j}\right) \\
&= O(1)\Bigg((\mathrm{E}\, X_1^2)^3 + \mathrm{E}\, X_1^2\ \mathrm{E}\, X_1^4 \sum_{p=1}^{n} O^4(l,p) \\
&\qquad + \left(\mathrm{E}\, X_1^3\right)^2 \left(\sum_{p=1}^{n} O^3(l,p)\right)^2 + \mathrm{E}\, X_1^6 \sum_{p=1}^{n} O^6(l,p)\Bigg) \\
&= O(1).
\end{aligned}$$

Together with (A.11), this yields the Lyapunov condition

$$\frac{1}{\left(\operatorname{var}\sum_{l=1}^{n} c^2_{\boldsymbol{Y}}(l)\right)^{3/2}} \sum_{l=1}^{n} \mathrm{E}\left|c^2_{\boldsymbol{Y}}(l) - \mathrm{E}\, c^2_{\boldsymbol{Y}}(l)\right|^3 = O(n^{-1/2}) \to 0,$$

which concludes the proof. ∎

Proof of Corollary 2.4. As convergence in Mallows' metric is equivalent to weak convergence and convergence of the first two moments (compare Bickel and Freedman (1981), Lemma 8.3), the claimed result follows immediately from Theorem 2.3. ∎

Proof of Theorem 2.5. By (A.2), (A.9) and due to $\mathrm{E}\, d^2_{\boldsymbol{Y},k}(l) = \mathrm{E}\, d^2_{\boldsymbol{X},k}(l)$ by construction, we get $\mathrm{E}\, T_n(\boldsymbol{Y}) = \mathrm{E}\, T_n(\boldsymbol{X})$. Further, as $\{X_t\}$ is 1-dependent and Gaussian, we get from Lemma A.2 (a) that

$$\begin{aligned}
&\operatorname{cov}(d^2_{\boldsymbol{X},k_1}(l_1), d^2_{\boldsymbol{X},k_2}(l_2)) \\
&= 2\delta_{k_1,k_2}\delta_{l_1,l_2}\gamma^2_{\boldsymbol{X}}(0) \\
&\quad + 4\delta_{k_1,k_2}\delta_{l_1,l_2}\left(\sum_{t=1}^{n-1}\left[h_{k_1,l_1}(t)h_{k_2,l_2}(t+1) + h_{k_1,l_1}(t+1)h_{k_2,l_2}(t)\right]\right)\gamma_{\boldsymbol{X}}(0)\gamma_{\boldsymbol{X}}(1) \\
&\quad + 2\left(\sum_{t=1}^{n-1}\left[h_{k_1,l_1}(t)h_{k_2,l_2}(t+1) + h_{k_1,l_1}(t+1)h_{k_2,l_2}(t)\right]\right)^2 \gamma^2_{\boldsymbol{X}}(1).
\end{aligned} \tag{A.12}$$

holds, such that using (A.9) and comparing the coefficients with (2.17) we get

$$\alpha_n = \frac{8}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\left(\sum_{t=1}^{n-1} h_{k,l}(t)h_{k,l}(t+1)\right) \tag{A.13}$$

$$\beta_n = \frac{2}{n}\sum_{k_1,k_2=1}^{D}\sum_{l_1=1}^{n_{k_1}}\sum_{l_2=1}^{n_{k_2}}\left(\sum_{t=1}^{n-1}[h_{k_1,l_1}(t)h_{k_2,l_2}(t+1)+h_{k_1,l_1}(t+1)h_{k_2,l_2}(t)]\right)^2 . \tag{A.14}$$

Similarly, from $\operatorname{cov}(d^2_{\boldsymbol{Y},k_1}(l_1), d^2_{\boldsymbol{Y},k_2}(l_2)) = 0$ if $k_1 \neq k_2$ or $l_1 \neq l_2$ and $\operatorname{var}(d^2_{\boldsymbol{Y},k}(l)) = \operatorname{var}(d^2_{\boldsymbol{X},k}(l))$ by construction, we get from Lemma A.2 (a)

$$\begin{aligned} n\operatorname{var}(T_n(\boldsymbol{Y})) &= \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\operatorname{var}(d^2_{\boldsymbol{X},k}(l)) = \frac{2}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\left\{\gamma_{\boldsymbol{X}}(0) + 2\sum_{t=1}^{n-1} h_{k,l}(t)h_{k,l}(t+1)\gamma_{\boldsymbol{X}}(1)\right\}^2 \\ &= (2-2/n)\gamma^2_{\boldsymbol{X}}(0) + \alpha_n\gamma_{\boldsymbol{X}}(0)\gamma_{\boldsymbol{X}}(1) + \beta_{\boldsymbol{W}_n}\gamma^2_{\boldsymbol{X}}(1) \end{aligned}$$

holds, where

$$\begin{aligned} \beta_{\boldsymbol{W}_n} &= \frac{2}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\left(\sum_{t=1}^{n-1}[h_{k,l}(t)h_{k,l}(t+1)+h_{k,l}(t+1)h_{k,l}(t)]\right)^2 \\ &= \frac{8}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\left(\sum_{t=1}^{n-1} h_{k,l}(t)h_{k,l}(t+1)\right)^2 . \end{aligned} \tag{A.15}$$

Since all summands of $\beta_{\boldsymbol{W}_n}$ are also summands of β_n, the difference $\beta_n - \beta_{\boldsymbol{W}_n}$ consists only of quadratic (i.e. nonnegative) summands. Consequently, we get a lower bound for the difference by choosing a suitable subset of those terms. To this end consider for $k=1$ the pairs (l_1, l_1+m), $m = M/2$, where the shifting is chosen such that for every index t either $h_{1,l_1}(t) = 0$ or $h_{1,l_1+m}(t+1) = 0$ with the exception of one $t_{l_1,M}$ for which $h_{1,l_1}(t) = h_0 \neq 0$ and $h_{1,l_1}(t+1) = h_{M-1} \neq 0$. Similarly, we also keep the terms corresponding to (l_2+m, l_2). Consequently, we obtain for $n \geqslant 2M$

$$\begin{aligned} \beta_n - \beta_{\boldsymbol{W}_n} &\geqslant \frac{2}{n}\sum_{k=1}^{D}\sum_{\substack{l_1,l_2=1\\ l_1\neq l_2}}^{n_k}\left(\sum_{t=1}^{n-1}[h_{k,l_1}(t)h_{k,l_2}(t+1)+h_{k,l_1}(t+1)h_{k,l_2}(t)]\right)^2 \\ &\geqslant \frac{4}{n}\sum_{l=1}^{n/2-m}\left(\sum_{t=1}^{n-1} h_{1,l}(t)h_{1,l+m}(t+1)\right)^2 = 2(h_0h_{M-1})^2(1-M/n) \geqslant (h_0h_{M-1})^2, \end{aligned}$$

proving (a). Now, let $\boldsymbol{W}_n = \boldsymbol{W}_{n,H}$ be the Haar-wavelet transform. Then, we obtain from (2.14) that $\sum_{t=1}^{n-1} h_{k,l}(t)h_{k,l}(t+1) = 1 - 3/2^k$, which leads together with (A.15) to

$$\beta_{\boldsymbol{W}_n} = \frac{8}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\left(1-\frac{3}{2^k}\right)^2 = 8\sum_{k=1}^{D}\frac{1}{2^k}\left(1-\frac{3}{2^k}\right)^2 \to \frac{16}{7} \tag{A.16}$$

as $n \to \infty$. This completes the proof of part (b). ■

Proof of Corollary 2.6. As convergence in Mallows' metric is equivalent to weak convergence and convergence of the first two moments (compare Bickel and Freedman (1981), Lemma 8.3), the claimed result follows immediately from Theorem 2.5. ■

Proof of Theorem 2.7. Denote

$$H_j(k_1,l_1,k_2,l_2) = \sum_{t=1}^{n-j} \left[h_{k_1,l_1}(t)\, h_{k_2,l_2}(t+j) + h_{k_1,l_1}(t+j)\, h_{k_2,l_2}(t) \right].$$

On the one hand, by an application of Lemma A.2 (b) and the Gaussianity, we get in the Wavelet domain

$$\begin{aligned}
n\,\mathrm{var}(T_n(\boldsymbol{X})) &= \frac{1}{n} \sum_{k_1,k_2=1}^{D} \sum_{l_1=1}^{n_{k_1}} \sum_{l_2=1}^{n_{k_2}} \mathrm{cov}(d^2_{\boldsymbol{X},k_1}(l_1), d^2_{\boldsymbol{X},k_2}(l_2)) \\
&= \frac{2}{n} \sum_{k_1,k_2=1}^{D} \sum_{l_1=1}^{n_{k_1}} \sum_{l_2=1}^{n_{k_2}} \mathrm{cov}^2(d_{\boldsymbol{X},k_1}(l_1), d_{\boldsymbol{X},k_2}(l_2)) \\
&= 2\gamma^2_{\boldsymbol{X}}(0)\, \frac{1}{n} \sum_{k=1}^{D} \sum_{l=1}^{n_k} \left(1 + 2 \sum_{j=1}^{n-1} a^j H_j(k,l,k,l) \right) \\
&\quad + 2\gamma^2_{\boldsymbol{X}}(0) \sum_{m=2}^{2n-2} a^m \sum_{j=\max(1,m-n+1)}^{\min(n-1,m-1)} \frac{1}{n} \sum_{k_1,k_2=1}^{D} \sum_{l_1=1}^{n_{k_1}} \sum_{l_2=1}^{n_{k_2}} H_j(k_1,l_1,k_2,l_2)\, H_{m-j}(k_1,l_1,k_2,l_2).
\end{aligned}$$

On the other hand, we get

$$\begin{aligned}
n\,\mathrm{var}(T_n(\boldsymbol{X})) &= n\,\mathrm{var}\left(\frac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2 \right) \\
&= n\,\mathrm{var}\left(\frac{1}{n} \sum_{t=1}^{n} (X_t - \mu)^2 \right) + n\,\mathrm{var}((\bar{X}_n - \mu)^2) - 2n\,\mathrm{cov}\left(\frac{1}{n} \sum_{t=1}^{n} (X_t - \mu)^2, (\bar{X}_n - \mu)^2 \right)
\end{aligned}$$

By expanding the latter three terms and exploiting the assumed Gaussianity, we can easily show that

$$n\,\mathrm{var}(T_n(\boldsymbol{X})) = 2\gamma^2_X(0)\xi_{0,n} + 2\gamma^2_X(0) \sum_{m\geqslant 1} \xi_{m,n} a^m,$$

holds with $|\xi_{m,n}| \leqslant C$ for some constant C that does not depend on m or n. Since both representations above hold for all $|a| < 1$, a comparison of coefficients yields

$$\begin{aligned}
&n\,\mathrm{var}(T_n(\boldsymbol{X})) \\
&= 2\gamma^2_{\boldsymbol{X}}(0)\, \frac{1}{n} \sum_{k=1}^{D} \sum_{l=1}^{n_k} \left(1 + 2 \sum_{j=1}^{n-1} a^j H_j(k,l,k,l) \right) + 2\gamma^2_{\boldsymbol{X}}(0) \sum_{m=2}^{2n-2} \xi_{m,n} a^m.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&n\,\mathrm{var}(T_n(\boldsymbol{Y})) \\
&= 2\gamma^2_{\boldsymbol{X}}(0)\, \frac{1}{n} \sum_{k=1}^{D} \sum_{l=1}^{n_k} \left(1 + 2 \sum_{j=1}^{n-1} a^j H_j(k,l,k,l) \right) \\
&\quad + 2\gamma^2_{\boldsymbol{X}}(0) \sum_{m=2}^{2n-2} a^m \sum_{j=\max(1,m-n+1)}^{\min(n-1,m-1)} \frac{1}{n} \sum_{k=1}^{D} \sum_{l=1}^{n_k} H_j(k,l,k,l)\, H_{m-j}(k,l,k,l).
\end{aligned}$$

Consequently,

$$\begin{aligned}
&n \operatorname{var}(T_n(\boldsymbol{X})) - n \operatorname{var}(T_n(\boldsymbol{Y}))\\
&= \gamma_{\boldsymbol{X}}^2(0) \sum_{m=2}^{2n-2} a^m \left(2\xi_{m,n} - \sum_{j=\max(1,m-n+1)}^{\min(n-1,m-1)} \frac{2}{n} \sum_{k=1}^{D} \sum_{l=1}^{n_k} H_j(k,l,k,l)\, H_{m-j}(k,l,k,l) \right)\\
&=: \gamma_{\boldsymbol{X}}^2(0) \sum_{m=2}^{2n-2} a^m \left(2\xi_{m,n} - \sum_{j=\max(1,m-n+1)}^{\min(n-1,m-1)} S_{n,m}(j) \right) =: \gamma_{\boldsymbol{X}}^2(0) \sum_{m\geqslant 2} a^m c_{m,n}.
\end{aligned}$$

By an application of the Cauchy-Schwarz inequality $|H_j(k_1,l_1,k_2,l_2)| \leqslant 2$, we get $S_{n,m}(j) = O(1)$ uniformly in m and n. Hence,

$$\sup_n |c_{m,n}| \leqslant Cm$$

for some constant $C > 0$. Using the boundedness of $c_{m,n}$ and a diagonal argument, one finds a subsequence $\alpha(n)$ such that $c_{m,\alpha(n)} \to c_m$ as $n \to \infty$. By an application of dominated convergence as ma^m is summable, we get

$$\sum_{m\geqslant 2} c_{m,\alpha(n)} a^m \to \sum_{m\geqslant 2} c_m a^m.$$

Furthermore with the notation of Theorem 2.5

$$\lim_{n\to\infty} |c_{2,\alpha(n)}| = |\beta_{\alpha(n)} - \beta_{\boldsymbol{W}_{\alpha(n)}}| \geqslant c > 0.$$

Consequently, $\sum_{m\geqslant 2} c_m a^m \neq 0$ (as a function of a), which gives the assertion. ■

Proof of Theorem 2.8.
Similar to the proof of Theorem 2.5, we get $\mathrm{E}\, T_n(\widetilde{\boldsymbol{Y}}) = \mathrm{E}\, T_n(\boldsymbol{X})$. Further, by construction, we have $\operatorname{cov}(d^2_{\widetilde{\boldsymbol{Y}},k_1}(l_1), d^2_{\widetilde{\boldsymbol{Y}},k_2}(l_2)) = 0$ if $k_1 \neq k_2$ and $\operatorname{cov}(d^2_{\widetilde{\boldsymbol{Y}},k}(l_1), d^2_{\widetilde{\boldsymbol{Y}},k}(l_2)) = \operatorname{cov}(d^2_{\boldsymbol{X},k}(l_1), d^2_{\boldsymbol{X},k}(l_2))$. By Lemma A.2 (a) we get

$$\begin{aligned}
n \operatorname{var}(T_n(\widetilde{\boldsymbol{Y}})) &= \frac{1}{n} \sum_{k=1}^{D} \sum_{l_1,l_2=1}^{n_k} \operatorname{cov}(d^2_{\boldsymbol{X},k}(l_1), d^2_{\boldsymbol{X},k}(l_2))\\
&= (2 - 2/n)\gamma_{\boldsymbol{X}}^2(0) + \alpha_n \gamma_{\boldsymbol{X}}(0)\gamma_{\boldsymbol{X}}(1) + \widetilde{\beta}_{\boldsymbol{W}_n} \gamma_{\boldsymbol{X}}^2(1),
\end{aligned}$$

where α_n as in (A.13) and

$$\widetilde{\beta}_{\boldsymbol{W}_n} = \frac{2}{n} \sum_{k=1}^{D} \sum_{l_1,l_2=1}^{n_k} \left(\sum_{t=1}^{n-1} [h_{k,l_1}(t) h_{k,l_2}(t+1) + h_{k,l_1}(t+1) h_{k,l_2}(t)] \right)^2. \tag{A.17}$$

Similarly, as in the proof of Theorem 2.5 $\beta_n - \widetilde{\beta}_{\boldsymbol{W}_n}$ consists only of positive summands and can be bounded from below by a suitable subset of summands. We choose $k_1 = 1$, $k_2 = 2$, and consider all pairs where $h_{2,l_2}(t) = h_0^{(2)} \neq 0$ while $h_{1,l_1}(t+1) = h_{M-1} \neq 0$, $t = 2, \ldots, n$, with $h_0^{(2)}$ as in (2.8). As the translation in the second finest level is by 4 while it is by 2 in the finest scale, we need to match properly $(1, l_1)$ with $(2, l_2)$, which is achieved for $l_1 = (m+2l_2) \mod n/2$, $l_2 = 1,\ldots,n/4-1$. There are only $n/4-1$ instead of $n/4$ matches because for $l_2 = n/4$ the element $h_0^{(2)}$ is on position n of $\boldsymbol{h}_{2,n/4}$. Similarly, we consider $h_{1,l_1}(t) = h_0$ and $h_{2,l_2}(t+1) = h_{M_2-1}^{(2)}$. There exist $n/4$ such matches if m is odd because in this case $M_2/4$ is an integer and $\boldsymbol{h}_{2,M_2/4}$ has the entry $h_{M_2-1}^{(2)}$ on position 1, while this can never happen for m even resulting in $n/4-1$ matches for $m/4$ even. Consequently, we obtain for $n \geqslant 4M$

$$\beta_n - \widetilde{\beta}_{\boldsymbol{W}_n} = \frac{2}{n}\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{D}\sum_{l_1=1}^{n_{k_1}}\sum_{l_2=1}^{n_{k_2}}\left(\sum_{t=1}^{n-1}[h_{k_1,l_1}(t)h_{k_2,l_2}(t+1)+h_{k_1,l_1}(t+1)h_{k_2,l_2}(t)]\right)^2$$

$$\geqslant \left(\frac{1}{2}-\frac{2}{n}\right)\left(\left(h_0^{(2)}h_{M-1}\right)^2+\left(h_{3M-3}^{(2)}h_0\right)^2\right) \geqslant \frac{1}{4}\left((h_0^{(2)}h_{M-1})^2+(h_0h_{3M-3}^{(2)})^2\right)$$

proving (a).

For (b), it holds for the Haar basis $\sum_{t=1}^{n-1} h_{k,l_1}(t)h_{k,l_2}(t+1) = -1/2^k$ for $l_2 = l_1+1$ as well as $\sum_{t=1}^{n-1} h_{k,l_1}(t+1)h_{k,l_2}(t) = -1/2^k$ for $l_2 = l_1-1$. All other sums are 0. Consequently,

$$\widetilde{\beta}_{\boldsymbol{W}_n} - \beta_{\boldsymbol{W}_n} = \frac{2}{n}\sum_{k=1}^{D}\sum_{\substack{l_1,l_2=1\\l_1\neq l_2}}^{n_k}\left(\sum_{t=1}^{n-1}[h_{k,l_1}(t)h_{k,l_2}(t+1)+h_{k,l_1}(t+1)h_{k,l_2}(t)]\right)^2$$

$$= \frac{4}{n}\sum_{k=1}^{D}(n_k-1)\left(-\frac{1}{2^k}\right)^2 = 4\sum_{k=1}^{D}\frac{1}{2^{3k}} - \frac{4}{n}\sum_{k=1}^{D}\frac{1}{2^{2k}} \to \frac{4}{7}$$

as $n\to\infty$. ■

Proof of Corollary 2.9. As convergence in Mallows' metric is equivalent to weak convergence and convergence of the first two moments (compare Bickel and Freedman (1981), Lemma 8.3), the claimed result follows immediately from Theorem 2.8. ■

Proof of Theorem 2.10. First, note that $h_{k,l}$ has only $M_k = (2^k-1)(M-1)+1$ consecutive non-zero elements, which are circularly shifted by 2^k. Consequently, for k_1 and k_2 and l_1 (or l_2) fixed $H_j(k_1,l_1,k_2,l_2) = 0$ except for $O(2^{|k_1-k_2|})$ values of l_2 (or l_1), where the constants do not depend on j. In particular, we get

$$\sum_{l_1=1}^{n_{k_1}}\sum_{l_2=1}^{n_{k_2}}|H_j(k_1,l_1,k_2,l_2)\,H_{m-j}(k_1,l_1,k_2,l_2)| = O\left(2^{|k_1-k_2|}\min(n_{k_1},n_{k_2})\right). \quad \text{(A.18)}$$

The proof of Theorem 2.10 is analogous to the proof of Theorem 2.7 with the exception that

$$n\operatorname{var}(T_n(\boldsymbol{X})) - n\operatorname{var}(T_n(\widetilde{\boldsymbol{Y}})) =: \gamma_{\boldsymbol{X}}^2(0)\sum_{m\geqslant 2}a^m\widetilde{c}_{m,n}$$

with $\sup_n|\widetilde{c}_{m,n}| \leqslant \widetilde{C}m$ for some constant $\widetilde{C}>0$, where by (A.18)

$$\frac{1}{n}\sum_{k}^{D}\sum_{l_1=1}^{n_k}\sum_{l_2=1}^{n_k}|H_j(k,l_1,k,l_2)\,H_{m-j}(k,l_1,k,l_2)| = O(1)\,\frac{1}{n}\sum_{k=1}^{D}n_k = O(1),$$

has been used. With the notation of Theorem 2.8, this leads to

$$|\widetilde{c}_{2,n}| = |\beta_n - \widetilde{\beta}_{W_n}| \geqslant \widetilde{c} > 0.$$

■

Proof of Theorem 2.11. From $\boldsymbol{X} = \boldsymbol{W}^T \boldsymbol{D}_{\boldsymbol{X}}$ we get

$$X_t = \sum_{k=1}^{D} \sum_{l=1}^{n_k} h_{k,l}(t) d_{\boldsymbol{X},k}(l) + \bar{X}_n$$

as $\frac{1}{\sqrt{n}} \upsilon_{\boldsymbol{X},D}(1) = \bar{X}_n$. Exploiting symmetry properties, this leads to

$$\begin{aligned}
\widehat{\gamma}_{\boldsymbol{X}}(1) &= \frac{1}{n} \sum_{t=1}^{n-1} (X_{t+1} - \bar{X}_n)(X_t - \bar{X}_n) \\
&= \frac{1}{n} \sum_{k_1=1}^{D} \sum_{l_1=1}^{n_{k_1}} \sum_{k_2=1}^{D} \sum_{l_2=1}^{n_{k_2}} \left(\sum_{t=1}^{n-1} h_{k_1,l_1}(t+1) h_{k_2,l_2}(t) \right) d_{\boldsymbol{X},k_1}(l_1) d_{\boldsymbol{X},k_2}(l_2) \\
&= \frac{1}{2n} \sum_{k_1=1}^{D} \sum_{l_1=1}^{n_{k_1}} \sum_{k_2=1}^{D} \sum_{l_2=1}^{n_{k_2}} \left(\sum_{t=1}^{n-1} [h_{k_1,l_1}(t+1) h_{k_2,l_2}(t) + h_{k_1,l_1}(t) h_{k_2,l_2}(t+1)] \right) d_{\boldsymbol{X},k_1}(l_1) d_{\boldsymbol{X},k_2}(l_2) \\
&= \frac{1}{2n} \sum_{k_1=1}^{D} \sum_{l_1=1}^{n_{k_1}} \sum_{k_2=1}^{D} \sum_{l_2=1}^{n_{k_2}} H_1(k_1, l_1, k_2, l_2)\, d_{\boldsymbol{X},k_1}(l_1) d_{\boldsymbol{X},k_2}(l_2) \qquad \text{(A.19)}
\end{aligned}$$

with analogous expressions for $\widehat{\gamma}_{\boldsymbol{Y}}(1)$ and $\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)$.

(a) As in the proof of Theorem 2.5, we get by construction of $\boldsymbol{Y}$ and from Lemma A.2 (a) that $\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1) = \frac{1}{4}(\beta_n - \beta_{\boldsymbol{W}_n})\gamma_{\boldsymbol{X}}(1)$ such that

$$|\,\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1)| \geqslant \frac{1}{4}(h_0 h_{M-1})^2 |\gamma_{\boldsymbol{X}}(1)| > 0.$$

As $\beta_n \to 4$ and $\beta_{\boldsymbol{W}_{n,H}} \to \frac{16}{7}$ for the Haar basis, we get $\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1) \to \frac{3}{7}\gamma_{\boldsymbol{X}}(1)$, such that $\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1) \to \frac{4}{7}\gamma_{\boldsymbol{X}}(1)$.

(b) From the proof of Theorem 2.8, we get similar as above

$$|\,\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)| = \frac{1}{4}(\beta_n - \widetilde{\beta}_{\boldsymbol{W}_n})|\gamma_{\boldsymbol{X}}(1)| \geqslant \frac{1}{16}\left((h_0^{(2)} h_{M-1})^2 + (h_0 h_{3M-3}^{(2)})^2 \right) |\gamma_{\boldsymbol{X}}(1)| > 0.$$

As $\widetilde{\beta}_{\boldsymbol{W}_{n,H}} \to \frac{20}{7}$ for the Haar basis, we get $\mathrm{E}\,\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1) \to \frac{5}{7}\gamma_{\boldsymbol{X}}(1)$. ■

Lemma A.3. *Under the assumptions of Theorem 2.5 and Theorem 2.8, respectively, we have (a)* $\widehat{\gamma}_{\boldsymbol{Y}}(1) \xrightarrow{P} \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1)$ *and (b)* $\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1) \xrightarrow{P} \mathrm{E}\,\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)$.

Proof. By the definition of $\boldsymbol{Y}$, Lemma A.2 (a) and the Gaussianity we get

$$\begin{aligned}
&\operatorname{cov}\left(d_{\boldsymbol{Y},k_1}(l_1) d_{\boldsymbol{Y},k_2}(l_2)\,,\, d_{\boldsymbol{Y},k_3}(l_3) d_{\boldsymbol{Y},k_4}(l_4)\right) \\
&= \begin{cases} O(1), & k_1 = k_3, l_1 = l_3, k_2 = k_4, l_2 = l_4 \quad \text{or} \quad k_1 = k_4, l_1 = l_4, k_2 = k_3, l_2 = l_3, \\ 0, & \text{else.} \end{cases}
\end{aligned}$$

By $H_1(k_1, l_1, k_2, l_2) = H_1(k_2, l_2, k_1, l_1)$, an application of (A.18) as well as (A.19) gives

$$\begin{aligned}
\operatorname{var}(\widehat{\gamma}_{\boldsymbol{Y}}(1)) &= O(1)\frac{1}{n^2}\sum_{k_1,k_2=1}^{D}\sum_{l_1=1}^{n_{k_1}}\sum_{l_2=1}^{n_{k_2}} H_1^2(k_1, l_1, k_2, l_2) \\
&= O(1)\frac{1}{n^2}\sum_{k_1,k_2=1}^{D} 2^{|k_1-k_2|}\min(n_{k_1}, n_{k_2}) = O(1)\frac{1}{n}\sum_{k_1=1}^{D}\sum_{k_2=1}^{D} 2^{-\min(k_1,k_2)} = O\left(\frac{\log n}{n}\right) = o(1),
\end{aligned}$$

proving (a).

Similarly, we get

$$\begin{aligned}
&\operatorname{cov}\left(d_{\widetilde{\boldsymbol{Y}},k_1}(l_1)d_{\widetilde{\boldsymbol{Y}},k_2}(l_2)\,,\,d_{\widetilde{\boldsymbol{Y}},k_3}(l_3)d_{\widetilde{\boldsymbol{Y}},k_4}(l_4)\right) \\
&= \begin{cases}
\mathrm{E}(d_{\boldsymbol{X},k_1}(l_1)\,d_{\boldsymbol{X},k_1}(l_3))\,\mathrm{E}(d_{\boldsymbol{X},k_2}(l_2)\,d_{\boldsymbol{X},k_2}(l_4)), & k_1 = k_3 \neq k_2 = k_4, \\
\mathrm{E}(d_{\boldsymbol{X},k_1}(l_1)\,d_{\boldsymbol{X},k_1}(l_4))\,\mathrm{E}(d_{\boldsymbol{X},k_2}(l_2)\,d_{\boldsymbol{X},k_2}(l_3)) & k_1 = k_4 \neq k_2 = k_3, \\
\operatorname{cov}\left(d_{\widetilde{\boldsymbol{X}},k_1}(l_1)\,d_{\widetilde{\boldsymbol{X}},k_1}(l_2), d_{\widetilde{\boldsymbol{X}},k_1}(l_3)\,d_{\widetilde{\boldsymbol{X}},k_1}(l_4)\right) & k_1 = k_2 = k_3 = k_4 \\
0, & \text{else.}
\end{cases}
\end{aligned}$$

For $k_1 = k_2 = k_3 = k_4$ we get by Gaussianity

$$\begin{aligned}
&\operatorname{cov}\left(d_{\widetilde{\boldsymbol{X}},k_1}(l_1)\,d_{\widetilde{\boldsymbol{X}},k_1}(l_2), d_{\widetilde{\boldsymbol{X}},k_1}(l_3)\,d_{\widetilde{\boldsymbol{X}},k_1}(l_4)\right) \\
&= \mathrm{E}(d_{\boldsymbol{X},k_1}(l_1)\,d_{\boldsymbol{X},k_1}(l_3))\,\mathrm{E}(d_{\boldsymbol{X},k_1}(l_2)\,d_{\boldsymbol{X},k_1}(l_4)) + \mathrm{E}(d_{\boldsymbol{X},k_1}(l_1)\,d_{\boldsymbol{X},k_1}(l_4))\,\mathrm{E}(d_{\boldsymbol{X},k_1}(l_2)\,d_{\boldsymbol{X},k_1}(l_3)).
\end{aligned}$$

Hence (A.19) yields

$$\begin{aligned}
&\operatorname{var}(\widehat{\gamma}_{\widetilde{\boldsymbol{Y}}}(1)) \\
&= \frac{1}{n^2}\sum_{k_1=1}^{D}\sum_{l_1,l_3=1}^{n_{k_1}}\sum_{k_2=1}^{D}\sum_{l_2,l_4=1}^{n_{k_2}} H_1(k_1, l_1, k_2, l_2)H_1(k_1, l_3, k_2, l_4) \\
&\qquad\qquad \times \mathrm{E}\left(d_{\boldsymbol{X},k_1}(l_1)d_{\boldsymbol{X},k_1}(l_3)\right)\mathrm{E}\left(d_{\boldsymbol{X},k_2}(l_2)d_{\boldsymbol{X},k_2}(l_4)\right) \\
&+ \frac{1}{n^2}\sum_{k_1=1}^{D}\sum_{l_1,l_4=1}^{n_{k_1}}\sum_{k_2=1}^{D}\sum_{l_2,l_3=1}^{n_{k_2}} H_1(k_1, l_1, k_2, l_2)H_1(k_2, l_3, k_1, l_4) \\
&\qquad\qquad \times \mathrm{E}\left(d_{\boldsymbol{X},k_1}(l_1)d_{\boldsymbol{X},k_1}(l_4)\right)\mathrm{E}\left(d_{\boldsymbol{X},k_2}(l_2)d_{\boldsymbol{X},k_2}(l_3)\right) \\
&= A + B.
\end{aligned}$$

By Lemma A.2 (a) we further get

$$\begin{aligned}
&\mathrm{E}\left(d_{\boldsymbol{X},k_1}(l_1)d_{\boldsymbol{X},k_1}(l_3)\right)\mathrm{E}\left(d_{\boldsymbol{X},k_2}(l_2)d_{\boldsymbol{X},k_2}(l_4)\right) \\
&= O(1)\delta_{l_1,l_3}\delta_{l_2,l_4} + O(1)\delta_{l_1,l_3}H_1(k_2, l_2, k_2, l_4) + O(1)\delta_{l_2,l_4}H_1(k_1, l_1, k_1, l_3) \\
&\quad + O(1)H_1(k_2, l_2, k_2, l_4)H_1(k_1, l_1, k_1, l_3)
\end{aligned}$$

leading to the decomposition $A = O(1)[A_1 + A_2 + A_3 + A_4]$ with (exemplary)

$$A_4 = \frac{1}{n^2} \sum_{k_1=1}^{D} \sum_{l_1,l_3=1}^{n_{k_1}} \sum_{k_2=1}^{D} \sum_{l_2,l_4=1}^{n_{k_2}} |H_1(k_1,l_1,k_2,l_2)H_1(k_1,l_3,k_2,l_4)H_1(k_2,l_2,k_2,l_4)H_1(k_1,l_1,k_1,l_3)|$$
$$\leqslant \frac{1}{n^2} \sum_{k_1,k_2=1}^{D} \sum_{l_1=1}^{n_{k_1}} \sum_{l_2=1}^{n_{k_2}} |H_1(k_1,l_1,k_2,l_2)| \sum_{l_3=1}^{n_{k_1}} |H_1(k_1,l_1,k_1,l_3)| \sum_{l_4=1}^{n_{k_2}} |H_1(k_2,l_2,k_2,l_4)|$$

where we used the fact that by the Cauchy-Schwarz inequality $|H_1(\cdot)| \leqslant 1$. Analogously to (A.18) we get

$$\sum_{l_1} |H_1(k_1,l_1,k_2,l_2)| = O\left(2^{|k_1-k_2|}\right), \qquad \sum_{l_2} |H_1(k_1,l_1,k_2,l_2)| = O\left(2^{|k_1-k_2|}\right),$$
$$\sum_{l_1=1}^{n_{k_1}} \sum_{l_2=1}^{n_{k_2}} |H_1(k_1,l_1,k_2,l_2)| = O\left(2^{|k_1-k_2|}\min(n_{k_1},n_{k_2})\right).$$

Consequently,

$$A_4 = O(1)\frac{1}{n^2} \sum_{k_1,k_2=1}^{D} 2^{|k_1-k_2|}\min(n_{k_1},n_{k_2}) = O\left(\frac{\log n}{n}\right)$$

as in the proof of (a). The terms A_1, A_2, A_3 can be dealt with similarly as can the term B after an analogous splitting step, completing the proof of (b). ■

Proof of Corollary 2.12. Since $\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) \to \gamma_{\boldsymbol{X}}(1)$, we get by Theorem 2.5 and Lemma A.3

$$\begin{aligned} &\rho_{\boldsymbol{X}}(1) - \widehat{\rho}_{\boldsymbol{Y}}(1) \\ &= \frac{\gamma_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1)}{\gamma_{\boldsymbol{X}}(0)} + \frac{\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1)}{\gamma_{\boldsymbol{X}}(0)} + \frac{\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1) - \widehat{\gamma}_{\boldsymbol{Y}}(1)}{\gamma_{\boldsymbol{X}}(0)} + \frac{\widehat{\gamma}_{\boldsymbol{Y}}(1)(\widehat{\gamma}_{\boldsymbol{Y}}(0) - \gamma_{\boldsymbol{X}}(0))}{\widehat{\gamma}_{\boldsymbol{Y}}(0)\gamma_{\boldsymbol{X}}(0)} \\ &= \frac{\mathrm{E}\,\widehat{\gamma}_{\boldsymbol{X}}(1) - \mathrm{E}\,\widehat{\gamma}_{\boldsymbol{Y}}(1)}{\gamma_{\boldsymbol{X}}(0)} + o_P(1), \end{aligned}$$

which together with Theorem 2.11 proves part (a)(i). Lemma A.3 and Theorem 2.11 give (a)(i). The arguments are analogous for (b) and therefore omitted. ■

B. Proofs of Section 3

Lemma B.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be Gaussian and $\gamma_{\boldsymbol{X}}(r) = O(|r|^{2d-1})$, $0 \leqslant d < 1/2$, as $|r| \to \infty$.*

(a) It holds,

(i) $\mathrm{var}(d^2_{\boldsymbol{X},k}(l)) = O(2^{4kd})$.

If additionally (3.5) *holds, then*

(ii) $\mathrm{cov}(d^2_{\boldsymbol{X},k_1}(l_1), d^2_{\boldsymbol{X},k_2}(l_2)) = O\left(2^{k_1}2^{k_2}2^{(4d-2)k_2}\right)$.

(b) *For $l_j \geqslant M-1$, $j=1,2$, and $|l_2 2^{k_2} - l_1 2^{k_1}| \geqslant M \max(2^{k_1}, 2^{k_2})$, it holds*

$$(i) \quad \operatorname{cov}(d^2_{\boldsymbol{X},k_1}(l_1), d^2_{\boldsymbol{X},k_1}(l_2)) = O\left(2^{4k_1 d}(|l_1 - l_2| - M)^{4d-2}\right).$$

If additionally (3.5) *holds, then we get for $k_1 \leqslant k_2$*

$$(ii) \quad \operatorname{cov}(d^2_{\boldsymbol{X},k_1}(l_1), d^2_{\boldsymbol{X},k_2}(l_2)) = O\left(2^{k_1} 2^{k_2} \left(|l_2 2^{k_2} - l_1 2^{k_1}| - M \max(2^{k_1}, 2^{k_2})\right)^{4d-2}\right).$$

(c) *For $l_1 < M-1$, $n_{k_2} - M 2^{k_1 - k_2} > l_2 \geqslant M-1$ and $|l_2 2^{k_2} - l_1 2^{k_1}| \geqslant M \max(2^{k_1}, 2^{k_2})$, it holds*

$$\begin{aligned}(i) \quad \operatorname{cov}(d^2_{\boldsymbol{X},k_1}(l_1), d^2_{\boldsymbol{X},k_2}(l_2)) &= O\left(2^{k_1} 2^{k_2} \left(|l_2 2^{k_2} - l_1 2^{k_1}| - M \max(2^{k_1}, 2^{k_2})\right)^{4d-2}\right) \\ &\quad + O\left(2^{4dk_1}(n_{k_1} - l_2 - M)^{4d-2}\right).\end{aligned}$$

If additionally (3.5) *holds, then*

$$\begin{aligned}(ii) \quad \operatorname{cov}(d_{\boldsymbol{X},k_1}(l_1), d_{\boldsymbol{X},k_2}(l_2)) &= O\left(2^{k_1} 2^{k_2} \left(|l_2 2^{k_2} - l_1 2^{k_1}| - M \max(2^{k_1}, 2^{k_2})\right)^{4d-2}\right) \\ &\quad + O\left(2^{k_1} 2^{k_2}(n - l_2 2^{k_2} - M 2^{k_1})^{4d-2}\right).\end{aligned}$$

(d) *It holds,*

$$\operatorname{var}(d_{\boldsymbol{X},k}(l)) = \begin{cases} \tau_k^2, & l \geqslant M-1 \\ \widetilde{r}_{kl}, & l < M-1 \end{cases} = \tau_k^2 + r_{kl} 1(l < M-1)$$

with $r_{kj} = \widetilde{r}_{kj} - \tau_k^2$ such that $\tau_k^2 = O(2^{2kd})$ as well as $\max_j |r_{kj}| = O(2^{2dk})$.

(e) *For $0 \leqslant d < 1/4$ it holds $\operatorname{var} T_n(\boldsymbol{X}) = O(1/n)$.*

All rates above are uniformly in l_1, l_2.

Proof. First, note that for $l_1, l_2 \geqslant M-1$ we get by (2.11)

$$\begin{aligned}\operatorname{cov}(d_{\boldsymbol{X},k_1}(l_1), d_{\boldsymbol{X},k_2}(l_2)) &= \sum_{s_1=0}^{M_{k_1}-1} h_{s_1}^{(k_1)} \sum_{s_2=0}^{M_{k_2}-1} h_{s_2}^{(k_2)} \gamma\left(l_2 2^{k_2} - s_2 - l_1 2^{k_1} + s_1\right) \\ &\leqslant \sum_{r=1-M_{k_1}}^{M_{k_2}-1} \left(\sum_{s=0}^{M_{k_1}-1} (h_{s-r}^{(k_2)})^2\right)^{1/2} \left|\gamma\left(l_2 2^{k_2} - l_1 2^{k_1} - r\right)\right| \\ &= O(1) \sum_{r=1-M_{k_1}}^{M_{k_2}-1} \left(\sum_{s=0}^{M_{k_1}-1} (h_{s-r}^{(k_2)})^2\right)^{1/2} \left|l_2 2^{k_2} - l_1 2^{k_1} - r\right|^{2d-1} \qquad (B.1)\end{aligned}$$

where

$$\sum_{s=0}^{M_{k_1}-1} (h_{s-r}^{(k_2)})^2 \leqslant 1, \qquad (B.2)$$

$$\sum_{r=1-M_{k_1}}^{M_{k_2}-1} \left|l_2 2^{k_2} - l_1 2^{k_1} - r\right|^{2d-1} = O\left(2^{2d\max(k_1,k_2)}\right). \qquad (B.3)$$

Due to Gaussianity, it holds

$$\operatorname{cov}(d^2_{\boldsymbol{X},k}(l_1), d^2_{\boldsymbol{X},k}(l_2)) = 2(\operatorname{cov}(d_{\boldsymbol{X},k}(l_1), d_{\boldsymbol{X},k}(l_2)))^2, \tag{B.4}$$

which implies (a)(i) if formula (2.11) holds. Replacing (B.2) by (3.5) yields (a)(ii) analogously. For (2.10), i.e. if circular wrapping is present, for one or both of the wavelet coefficients analogous bounds hold for both summands of (2.10), so that we arrive at (a)(i) as the covariance is a bilinear form and each of the possibly 4 terms is bounded in the same way.

By $|l_2 2^{k_2} - l_1 2^{k_1}| \geqslant M \max(2^{k_1}, 2^{k_2})$ we can replace (B.3) by

$$\sum_{r=1-M_{k_1}}^{M_{k_2}-1} \left| l_2 2^{k_2} - l_1 2^{k_1} - r \right|^{2d-1} = O\left(2^{\max(k_1,k_2)} \left(|l_2 2^{k_2} - l_1 2^{k_1}| - M \max(2^{k_1}, 2^{k_2}) \right)^{2d-1} \right). \tag{B.5}$$

Together with (B.4) this implies (b).

In the situation of (c) $d_{\boldsymbol{X},k_1}(l_1)$ is wrapped (i.e. follows (2.10)) while $d_{\boldsymbol{X},k_2}(l_2)$ is not wrapped (i.e. follows (2.11)). The assertions are obtained analogously by separately treating the two sums in $d_{\boldsymbol{X},k_1}(l_1)$ and noting that

$$n - l_2 2^{k_2} + l_1 2^{k_1} + r \geqslant n - l_2 2^{k_2} - M 2^{k_1}.$$

By stationarity of $\{X_t, t \in \mathbb{Z}\}$, we get that $\mathrm{E}(d^2_{\boldsymbol{X},k}(l))$ is identical for all $l \geqslant M-1$ and hence does not depend on l. Par (a)(i) implies $\mathrm{E}(d^2_{\boldsymbol{X},k}(l)) = O(2^{2dk})$, which yields (d). Assertion (e) is a straight-forward calculation using the rate of decay of the covariances by noting that due to Gaussianity $\operatorname{cov}(X_t^2, X_{t+r}^2) = O(|r|^{4d-2})$. ■

Proof of Theorem 3.1. Analogously to (A.2), we have $T_n(\boldsymbol{X}^*_{\boldsymbol{W}}) = \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k} d_k^{*2}(l)$, which implies

$$\mathrm{E}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}})) = \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k} \mathrm{E}^*(d_k^{*2}(l)) = \frac{1}{n}\sum_{k=1}^{D}\sum_{j=1}^{n_k} d^2_{\boldsymbol{X},k}(j) = T_n(\boldsymbol{X}).$$

For the conditional variance, as $\{d_k^*(l) : k, l\}$ are i.i.d. with $d_k^*(l)$ uniformly distributed on the set $\{d_k(1), \ldots, d_k(n_k)\}$, we obtain

$$\begin{aligned}
n \operatorname{var}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}})) &= \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k} \operatorname{var}^*\left(d_k^{*2}(l)\right) \\
&= \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\left\{ \frac{1}{n_k}\sum_{j=1}^{n_k} d^4_{\boldsymbol{X},k}(j) - \left(\frac{1}{n_k}\sum_{j=1}^{n_k} d^2_{\boldsymbol{X},k}(j)\right)^2 \right\} \\
&= \frac{1}{n}\sum_{k=1}^{D}\sum_{j=1}^{n_k} d^4_{\boldsymbol{X},k}(j) - \frac{1}{n^2}\sum_{k=1}^{D} 2^k \sum_{j_1,j_2=1}^{n_k} d^2_{\boldsymbol{X},k}(j_1) d^2_{\boldsymbol{X},k}(j_2).
\end{aligned}$$

Further, by taking unconditional expectations above and rearranging terms, we get

$$\begin{aligned}
&\mathrm{E}\left(n\,\mathrm{var}^*(T_n(\boldsymbol{X}^*_{\boldsymbol{W}}))\right) \\
&= \frac{1}{n}\sum_{k=1}^{D}\sum_{j=1}^{n_k}\mathrm{var}\left(d^2_{\boldsymbol{X},k}(j)\right) + \sum_{k=1}^{D}\left(\frac{1}{n}\sum_{j=1}^{n_k}\left(\mathrm{E}(d^2_{\boldsymbol{X},k}(j))\right)^2 - \frac{1}{2^k}\left(\frac{1}{n_k}\sum_{j=1}^{n_k}\mathrm{E}\left(d^2_{\boldsymbol{X},k}(j)\right)\right)^2\right) \\
&\qquad - \frac{1}{n^2}\sum_{k=1}^{D}2^k\sum_{j_1,j_2=1}^{n_k}\mathrm{cov}\left(d^2_{\boldsymbol{X},k}(j_1), d^2_{\boldsymbol{X},k}(j_2)\right) \\
&=: A_1 + A_2 - A_3.
\end{aligned}$$

As $\mathrm{var}(d^2_{\boldsymbol{X},k}(j)) = \mathrm{var}(d^2_{\boldsymbol{Y},k}(j))$ by construction of $\boldsymbol{Y}$, we get $A_1 = n\,\mathrm{var}(T_n(\boldsymbol{Y}))$.

First consider (b) with $d < 1/4$. ¿From Lemma B.1(d) and on noting that the mixed terms cancel, we get

$$\begin{aligned}
A_2 &= \frac{1}{n}\sum_{k=1}^{D}\sum_{j=1}^{n_k} r^2_{kj}1_{\{j<M-1\}} - \sum_{k=1}^{D}\frac{1}{2^k}\left(\frac{1}{n_k}\sum_{j=1}^{n_k} r_{kj}1_{\{j<M-1\}}\right)^2 \\
&= O\left(\frac{1}{n}\sum_{k=1}^{D}2^{4dk}\right) + O\left(\frac{1}{n^2}\sum_{k=1}^{D}2^{(4d+1)k}\right) = O(1)\begin{cases}\frac{\log n}{n}, & d=0,\\ n^{4d-1}, & d>0,\end{cases} \quad = o(1)
\end{aligned}$$

where $\sum_{k=1}^{D}2^{sk} = O(2^{sD}) = O(n^s)$ for $s > 0$ has been used. Concerning A_3, we can split the summation in the four parts (i) $j_1 \geqslant M-1$ as well as $|j_1 - j_2| \geqslant M$, (ii) $j_1 \geqslant M-1$ and $|j_1 - j_2| < M$, (iii) $j_1 < M-1$ and $j_2 < n_k - M$ and finally (iv) $j_1 < M-1$ and $j_2 > n_k - M$. We can then treat the summands in (i) by Lemma B.1 (b) the ones in (ii) and (iv) by (a) and the ones in (iii) by (c) to obtain for $d < 1/4$

$$\begin{aligned}
A_3 &= \frac{1}{n^2}\sum_{k=1}^{D}2^k\sum_{j_1,j_2=1}^{n_k}\mathrm{cov}\left(d^2_{\boldsymbol{X},k}(j_1), d^2_{\boldsymbol{X},k}(j_2)\right) \\
&= \frac{1}{n}\sum_{k}2^{4kd}\sum_{M<|r|<n_k}(|r|-M)^{4d-2} + \frac{1}{n}\sum_{k=1}^{D}2^{4kd} + \frac{1}{n^2}\sum_{k=1}^{D}2^{k+4dk}\sum_{j_2=1}^{n_k}(n_k - j_2 - M)^{4d-2} \\
&= O(n^{4d-1}) = o(1).
\end{aligned}$$

For $d = 1/4$, the different normalization immediately leads to $A_2/\log n = o(1)$ while for A_3 the sums over r are no longer $O(1)$ but only $O(D-k)$. However,

$$\frac{A_3}{\log n} = O(1)\frac{1}{n\log n}\sum_{k=1}^{D}(D-k)2^k = O(1)\frac{1}{n\log n}\sum_{k=0}^{D-1}k2^{D-k} = O(1)\frac{1}{\log n}\sum_{k=0}^{D-1}k2^{-k} = o(1).$$

This completes the proof for $d = 1/4$.

The assertion in a) holds because all rates from Lemma B.1 remain true in the i.i.d. case even for non-Gaussian data because all involved covariances are either bounded by one (in (a)) or are zero due to the independence (in (b) and (c)) which shows the negligibility of A_3. Lemma B.1 (d) also remains true with $r_{k,l} = 0$, which even shows that $A_2 = 0$ for i.i.d. data. ■

Proof of Theorem 3.2. We have $T_n(\widetilde{\boldsymbol{X}}^*_{\boldsymbol{W}}) = \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\widetilde{d}_k^{*2}(l)$ and by splitting the sum with respect to the (non-overlapping) bootstrap blocking, we get for the mean

$$\mathrm{E}^*(T_n(\widetilde{\boldsymbol{X}}^*_{\boldsymbol{W}})) = \frac{1}{n}\sum_{k=1}^{D}\sum_{s=1}^{N_k}\mathrm{E}^*\left(\sum_{j=(s-1)L_k+1}^{sL_k}\widetilde{d}_k^{*2}(j)\right) = \frac{1}{n}\sum_{k=1}^{D}\sum_{s=1}^{N_k}\frac{1}{N_k}\sum_{t=1}^{N_k}\sum_{j=(t-1)L_k+1}^{tL_k}d^2_{\boldsymbol{X},k}(j)$$
$$= \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}d^2_{\boldsymbol{X},k}(l) = T_n(\boldsymbol{X})$$

and as $\widetilde{d}^*_{k_1}(l_1)$ and $\widetilde{d}^*_{k_2}(l_2)$ are conditionally independent for $k_1 \neq k_2$, but also if l_1 and l_2 are not in the same block, we obtain for the variance

$$n\,\mathrm{var}^*(T_n(\widetilde{\boldsymbol{X}}^*_{\boldsymbol{W}})) = \frac{1}{n}\sum_{k=1}^{D}\sum_{s=1}^{N_k}\mathrm{var}^*\left(\sum_{j=(s-1)L_k+1}^{sL_K}\widetilde{d}_k^{*2}(j)\right)$$
$$= \frac{1}{n}\sum_{k=1}^{D}\sum_{s=1}^{N_k}\left\{\frac{1}{N_k}\sum_{t=1}^{N_k}\left(\sum_{j=(t-1)L_k+1}^{tL_k}d^2_{\boldsymbol{X},k}(j)\right)^2 - \left(\frac{1}{N_k}\sum_{t=1}^{N_k}\sum_{j=(t-1)L_k+1}^{tL_k}d^2_{\boldsymbol{X},k}(j)\right)^2\right\}$$
$$= \frac{1}{n}\sum_{k=1}^{D}\sum_{t=1}^{N_k}\sum_{j_1,j_2=(t-1)L_k+1}^{tL_k}d^2_{\boldsymbol{X},k}(j_1)d^2_{\boldsymbol{X},k}(j_2)$$
$$\quad - \frac{1}{n}\sum_{k=1}^{D}\frac{1}{N_k}\sum_{t_1,t_2=1}^{N_k}\sum_{j_1=(t_1-1)L_k+1}^{t_1L_k}\sum_{j_2=(t_2-1)L_k+1}^{t_2L_k}d^2_{\boldsymbol{X},k}(j_1)d^2_{\boldsymbol{X},k}(j_2).$$

Further, by taking unconditional expectations above and rearranging terms, we get

$$\mathrm{E}\left(n\,\mathrm{var}^*(T_n(\widetilde{\boldsymbol{X}}^*_{\boldsymbol{W}}))\right)$$
$$= \frac{1}{n}\sum_{k=1}^{D}\sum_{t=1}^{N_k}\sum_{j_1,j_2=(t-1)L_k+1}^{tL_k}\mathrm{cov}(d^2_{\boldsymbol{X},k}(j_1), d^2_{\boldsymbol{X},k}(j_2))$$
$$\quad + \frac{1}{n}\sum_{k=1}^{D}\left(\sum_{t=1}^{N_k}\left(\sum_{j=(t-1)L_k+1}^{tL_k}\mathrm{E}(d^2_{\boldsymbol{X},k}(j))\right)^2 - \frac{1}{N_k}\left(\sum_{l=1}^{n_k}\mathrm{E}(d^2_{\boldsymbol{X},k}(l))\right)^2\right)$$
$$\quad - \frac{1}{n}\sum_{k=1}^{D}\frac{1}{N_k}\sum_{l_1,l_2=1}^{n_k}\mathrm{cov}(d^2_{\boldsymbol{X},k}(l_1), d^2_{\boldsymbol{X},k}(l_2))$$
$$=: \widetilde{A}_1 + \widetilde{A}_2 - \widetilde{A}_3.$$

Before we treat the above terms let us first consider the following sums which will play an important role in the proofs: By $2^{b_k} \leqslant n_k = n/2^k$ we get for $d < 1/4$

$$\frac{1}{n}\sum_{k=1}^{D}2^{4dk+b_k} = \sum_{k\leqslant A(n)}2^{(4d-1)k}\,\frac{2^{b_k+k}}{n} + \sum_{k\geqslant A(n)}2^{(4d-1)k}\,\frac{2^{b_k}}{n_k},$$
$$\leqslant \sup_{k\leqslant A(n)}\frac{2^{b_k+k}}{n}\,\frac{1}{1-2^{4d-1}} + \sum_{k\geqslant A(n)}2^{(4d-1)k} = o(1). \tag{B.6}$$

Furthermore,

$$\sum_{k=1}^{D} b_k 2^{-k-b_k} \leqslant \sup_{k\leqslant A(n)} \frac{b_k}{2^{b_k}} \sum_{k=1}^{A(n)} 2^{-k} + \sum_{k>A(n)} 2^{-k} = o(1). \tag{B.7}$$

Similarly,

$$\begin{aligned}\sum_{k=1}^{D} 2^{(4d-1)(b_k+k)} &\leqslant \sup_{k\leqslant A(n)} 2^{(4d-1)b_k} \sum_{k\leqslant A(n)} 2^{(4d-1)k} + \sum_{k\geqslant A(n)} 2^{(4d-1)(k+b_k)} \\ &\leqslant \sup_{k\leqslant A(n)} 2^{(4d-1)b_k} \frac{1}{1-2^{4d-1}} + \sum_{k\geqslant A(n)} 2^{(4d-1)k} = o(1). \end{aligned} \tag{B.8}$$

Similar to the proof of Theorem 3.1, we can treat $\widetilde{A}_2$ with the help of Lemma B.1(d) to obtain

$$\begin{aligned}\widetilde{A}_2 &= \frac{2}{n}\sum_{k=1}^{D}\tau_k^2\sum_{t=1}^{N_k}\sum_{j=(t-1)L_k+1}^{tL_k} r_{kj}1(j<M-1) + \frac{1}{n}\sum_{k=1}^{D}\sum_{t=1}^{N_k}\left(\sum_{j=(t-1)L_k+1}^{tL_k} r_{kj}1(j<M-1)\right)^2 \\ &\quad - \frac{2}{n}\sum_{k=1}^{D}\frac{1}{N_k}\frac{n}{2^k}\tau_k^2\sum_{l=1}^{n_k} r_{kl}1_{\{l<M-1\}} - \frac{1}{n}\sum_{k=1}^{D}\frac{1}{N_k}\left(\sum_{l=1}^{n_k} r_{kl}1_{\{l<M-1\}}\right)^2 \\ &= O\left(\frac{1}{n}\sum_{k=1}^{D}2^{4dk}\right) + O\left(\frac{1}{n}\sum_{k=1}^{D}2^{4dk+b_k}\right) + O\left(\frac{1}{n^2}\sum_{k=1}^{D}2^{4dk+k+b_k}\right) \\ &= O\left(\frac{1}{n}\sum_{k=1}^{D}2^{4dk+b_k}\right) = o(1)\end{aligned}$$

by (B.6). Concerning $\widetilde{A}_3$, from Lemma B.1, we get similarly to the treatment of A_3 in the proof of Theorem 3.1

$$\begin{aligned}\widetilde{A}_3 &= \frac{1}{n}\sum_{k=1}^{D}\frac{1}{N_k}\sum_{l_1,l_2=1}^{n_k}\operatorname{cov}(d^2_{\mathbf{X},k}(l_1), d^2_{\mathbf{X},k}(l_2)) \\ &= O(1)\frac{1}{n}\sum_{k=1}^{D}2^{b_k+4kd}\sum_{M<|r|<n_k}(|r|-M)^{4d-2} + O(1)\frac{1}{n}\sum_{k=1}^{D}2^{b_k+4kd} \\ &\quad + O(1)\frac{1}{n^2}\sum_{k=1}^{D}2^{k+b_k+4kd}\sum_{l_2}(n_k-l_2-M)^{4d-2} + O(1)\frac{1}{n^2}\sum_{k=1}^{D}2^{k+b_k+4kd} \\ &= O(1)\frac{1}{n}\sum_{k=1}^{D}2^{b_k+4kd} = o(1)\end{aligned}$$

by (B.6). Now, it remains to show that $n\operatorname{var}(T_n(\widetilde{\boldsymbol{Y}})) - \widetilde{A}_1$ is of low order. By using the fact that

$$n\operatorname{var}(T_n(\widetilde{\boldsymbol{Y}})) = \frac{1}{n}\sum_{k=1}^{D}\sum_{t_1,t_2=1}^{N_k}\sum_{j_1=(t_1-1)L_k+1}^{t_1L_k}\sum_{j_2=(t_2-1)L_k+1}^{t_2L_k}\operatorname{cov}(d^2_{\mathbf{X},k}(j_1), d^2_{\mathbf{X},k}(j_2)),$$

by symmetry, this leads to

$$
\begin{aligned}
n\operatorname{var}(T_n(\widetilde{\boldsymbol{Y}})) - \widetilde{A}_1 &= \frac{2}{n}\sum_{k=1}^{D}\sum_{\substack{t_1,t_2=1\\ t_1<t_2}}^{N_k}\sum_{j_1=(t_1-1)L_k+1}^{t_1L_k}\sum_{j_2=(t_2-1)L_k+1}^{t_2L_k}\operatorname{cov}(d^2_{\boldsymbol{X},k}(j_1), d^2_{\boldsymbol{X},k}(j_2))\\
&\leqslant \widetilde{B}_1 + \widetilde{B}_2 + \widetilde{B}_3 + \widetilde{B}_4,
\end{aligned}
$$

where we use a similar splitting as in $\widetilde{A}_3$ above in order to use different parts of Lemma B.1: $\widetilde{B}_1$ is the sum over all indices fulfilling $j_2 - j_1 > M$ as well as $j_1 \geqslant \lceil M_k/2^k\rceil$. $\widetilde{B}_2$ is the sum over $j_2 - j_1 \leqslant M$. These two are the main parts which yield the restrictions on the bandwidth. The remaining terms have to be treated differently due to the wrapping effect present in wavelet coefficients: $\widetilde{B}_3$ is the sum over $j_1 < \lceil M_k/2^k\rceil, n_k - M > j_2 \geqslant \lceil M_k/2^k\rceil, j_2 - j_1 > M$ and $\widetilde{B}_4$ the sum over $j_1 \leqslant M$ as well as $j_2 \geqslant n_k - M$.

To treat $\widetilde{B}_1$, first note that for a general positive function $f(\cdot)$ it holds

$$
\begin{aligned}
\sum_{j_1=1}^{L}\sum_{j_2=L+1}^{2L} f(j_2-j_1) &= \sum_{j_1=1}^{L}\sum_{r=L+1-j_1}^{2L-j_1} f(r) = \sum_{r=1}^{2L-1} f(r) \sum_{j_1=\max(1,L+1-r)}^{\min(L,2L-r)}\\
&\leqslant \sum_{r=1}^{2L} r f(r), \qquad \text{(B.9)}
\end{aligned}
$$

as well as

$$
\sum_{j_2\geqslant j_1+L} f(j_2-j_1) \leqslant \sum_{r\geqslant L} f(r). \qquad \text{(B.10)}
$$

Combined with the results of Lemma B.1(b)(i) this allows for the following splitting of the sum

$$
\begin{aligned}
\widetilde{B}_1 &= \frac{1}{n}\sum_{k=1}^{D}\sum_{\substack{t_1,t_2=1\\ t_1<t_2}}^{N_k}\sum_{j_1=(t_1-1)L_k+1}^{t_1L_k}\sum_{j_2=(t_2-1)L_k+1}^{t_2L_k}\operatorname{cov}(d^2_{\boldsymbol{X},k}(j_1), d^2_{\boldsymbol{X},k}(j_2))1_{\{j_1\geqslant\lceil M_k/2^k\rceil, j_2-j_1>M\}}\\
&= O(1)\frac{1}{n}\sum_{k=1}^{D}2^{4dk}\sum_{t_1}\left(\sum_{|r|>M}^{L_k} r(r-M)^{4d-2} + L_k\sum_{r>\max(M,L_k)}|r-M|^{4d-2}\right)\\
&= O(1)\sum_{k=1}^{D}2^{(4d-1)k}2^{-b_k}\begin{cases}2^{4db_k}, & d\neq 0,\\ b_k, & d=0,\end{cases} + O(1)\sum_{k=1}^{D}2^{(4d-1)k}2^{b_k(4d-1)} = o(1)
\end{aligned}
$$

by (B.7) as well as (B.8). Concerning $\widetilde{B}_2$ we get

$$
\begin{aligned}
\widetilde{B}_2 &= \frac{1}{n}\sum_{k=1}^{D}\sum_{\substack{t_1,t_2=1\\ t_1<t_2}}^{N_k}\sum_{j_1=(t_1-1)L_k+1}^{t_1L_k}\sum_{j_2=(t_2-1)L_k+1}^{t_2L_k}\operatorname{cov}(d^2_{\boldsymbol{X},k}(j_1), d^2_{\boldsymbol{X},k}(j_2))1_{\{j_2-j_1\leqslant M\}}\\
&= O(1)\frac{1}{n}\sum_{k=1}^{D}2^{4dk}\sum_{t_1=1}^{N_k-1}\sum_{j_1=(t_1-1)L_k+1}^{t_1L_k}\sum_{j_2=t_1L_k+1}^{n/2^k}1_{\{j_2-j_1\leqslant M\}}\\
&= O(1)\frac{1}{n}\sum_{k=1}^{D}2^{4dk}N_k = O(1)\sum_{k=1}^{D}2^{(4d-1)k-b_k} = O(1)\sum_{k=1}^{D}2^{(4d-1)(b_k+k)} = o(1)
\end{aligned}
$$

by (B.8). Concerning $\widetilde{B}_3$ we need to apply Lemma B.1 (c) (i), where the first summand there can be treated analogously to $\widetilde{B}_1$, yielding

$$\begin{aligned}\widetilde{B}_3 &= o(1) + O(1)\,\frac{1}{n}\sum_{k=1}^{D}\sum_{t_2=1}^{N_k}\sum_{j_2=(t_2-1)L_k+1}^{t_2L_k}\Big(n/2^k - j_2 - M\Big)^{4d-2}2^{4dk}\\ &= o(1) + O(1)\,\frac{1}{n}\sum_{k=1}^{D}2^{4dk}\sum_{r\in\mathbb{Z}}r^{4d-2} = o(1) + O(n^{4d-1}) = o(1).\end{aligned}$$

In the term $\widetilde{B}_4$ the number of summands in each level is bounded by M^2, so that we get by Lemma B.1 (a) that $\widetilde{B}_4 = O(n^{4d-1}) = o(1)$. ∎

Proof of Theorem 3.3. By construction of the rectangular wavelet lattice and splitting the sum with respect to the (non-overlapping) bootstrap blocking, we can write

$$\begin{aligned}T_n(\check{\boldsymbol{X}}^*_{\boldsymbol{W}}) &= \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}\check{d}_k^{*2}(l) = \frac{1}{n}\sum_{k=1}^{D}\sum_{l=1}^{n_k}q_k^{*2}(2^{k-1}l) = \frac{1}{n}\sum_{k=1}^{D}\frac{1}{2^{k-1}}\sum_{l=1}^{n/2}q_k^{*2}(l)\\ &= \frac{1}{n}\sum_{l=1}^{n/2}\left(\sum_{k=1}^{D}2^{-k+1}q_k^{*2}(l)\right) = \frac{1}{n}\sum_{s=1}^{N}\left(\sum_{t=(s-1)L+1}^{sL}\sum_{k=1}^{D}2^{-k+1}q_k^{*2}(t)\right),\end{aligned} \tag{B.11}$$

where $N = n/(2L)$ and $L = 2^b$. Now, as $\{\sum_{t=(s-1)L+1}^{sL}\frac{1}{n}\sum_{k=1}^{D}2^{-k+1}q_k^{*2}(t), s = 1,\dots,N\}$ are i.i.d. conditionally on $\{q_k(t), k, t\}$, we get for the mean

$$\begin{aligned}\mathrm{E}^*(T_n(\check{\boldsymbol{X}}^*_{\boldsymbol{W}})) &= \frac{1}{n}\sum_{s=1}^{N}\mathrm{E}^*\left(\sum_{t=(s-1)L+1}^{sL}\sum_{k=1}^{D}2^{-k+1}q_k^{*2}(t)\right)\\ &= \frac{1}{n}\sum_{s=1}^{N}\frac{1}{N}\sum_{p=1}^{N}\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D}2^{-k+1}q_{\boldsymbol{X},k}^{2}(t) = T_n(\boldsymbol{X})\end{aligned}$$

by a representation of $T_n(\boldsymbol{X})$ analogous to (B.11). For the variance we get

$$\begin{aligned}n\,\mathrm{var}^*(T_n(\check{\boldsymbol{X}}^*_{\boldsymbol{W}})) &= \frac{N}{n}\,\mathrm{var}^*\left(\sum_{t_1=1}^{L}\sum_{k_1=1}^{D}2^{-k_1+1}q_{k_1}^{*2}(t_1)\right)\\ &= \frac{1}{n}\sum_{p=1}^{N}\left(\sum_{t_1=(p-1)L+1}^{pL}\sum_{k_1=1}^{D}2^{-k_1+1}q_{\boldsymbol{X},k_1}^{2}(t_1)\right)^2 - \frac{n}{N}T_n^2(\boldsymbol{X}).\end{aligned}$$

Taking unconditional expectations above and rearranging terms leads to

$$\begin{aligned}&\mathrm{E}\left(n\,\mathrm{var}^*(T_n(\check{\boldsymbol{X}}^*_{\boldsymbol{W}}))\right)\\ &= \frac{1}{n}\sum_{p=1}^{N}\mathrm{var}\left(\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D}2^{-k_1+1}q_{\boldsymbol{X},k}^{2}(t)\right)\\ &\quad+\left\{\frac{1}{n}\sum_{p=1}^{N}\left(\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D}2^{-k+1}\,\mathrm{E}\,q_{\boldsymbol{X},k}^{2}(t)\right)^2 - \frac{n}{N}(\mathrm{E}\,T_n(\boldsymbol{X}))^2\right\}\\ &\quad-\frac{n}{N}\,\mathrm{var}(T_n(\boldsymbol{X}))\\ &= \check{A}_1 + \check{A}_2 - \check{A}_3.\end{aligned}$$

We get from Lemma B.1 (e) (for $d < 1/4$).

$$\check{A}_3 = O\left(\frac{1}{N}\right) = O\left(\frac{L}{n}\right) = o(1).$$

Furthermore, by Lemma B.1 (d) we get

$$\begin{aligned}
\check{A}_2 &= \frac{1}{n}\sum_{p=1}^{N}\left(\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1}\,\mathrm{E}\,q^2_{\boldsymbol{X},k}(t)\right)^2 - \frac{n}{N}\left(\frac{1}{n}\sum_{p=1}^{N}\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1}\,\mathrm{E}\,q^2_{\boldsymbol{X},k}(t)\right)^2 \\
&= \frac{1}{n}\sum_{p=1}^{N}\left(\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1} O(2^{2kd}) 1_{\{t<2^k M\}}\right)^2 \\
&\quad - \frac{1}{Nn}\left(\sum_{p=1}^{N}\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1} O(2^{2kd}) 1_{\{t<2^k M\}}\right)^2 \\
&\quad + \frac{2}{n}\sum_{p=1}^{N}\left(\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1} O(2^{2kd}) 1_{\{t<2^k M\}}\right)\left(\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1}\tau_k^2\right) \\
&\quad - \frac{2}{Nn}\left(\sum_{p=1}^{N}\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1} O(2^{2kd}) 1_{\{t<2^k M\}}\right)\left(\sum_{p=1}^{N}\sum_{t=(p-1)L+1}^{pL}\sum_{k=1}^{D} 2^{-k+1}\tau_k^2\right) \\
&= \begin{cases} Ln^{4d-2}, & 0 < d < \frac{1}{4}, \\ L\log n/n, & d = 0, \end{cases} \qquad = o(1).
\end{aligned}$$

Further, by using the representation

$$n\,\mathrm{var}(T_n(\boldsymbol{X})) \;=\; \frac{1}{n}\sum_{k_1,k_2=1}^{D}\sum_{t_1,t_2=1}^{N}\sum_{j_1=(t_1-1)L+1}^{t_1L}\sum_{j_2=(t_2-1)L+1}^{t_1L}\frac{1}{2^{k_1-1}}\frac{1}{2^{k_2-1}}\,\mathrm{cov}(q^2_{\boldsymbol{X},k_1}(j_1), q^2_{\boldsymbol{X},k_2}(j_2))$$

we obtain

$$\begin{aligned}
n\,\mathrm{var}(T_n(\boldsymbol{X})) - \check{A}_1 &= \frac{1}{n}\sum_{k_1,k_2=1}^{D}\sum_{\substack{t_1,t_2=1\\ t_1\neq t_2}}^{N}\sum_{j_1=(t_1-1)L+1}^{t_1L}\sum_{j_2=(t_2-1)L+1}^{t_1L}\frac{1}{2^{k_1-1}}\frac{1}{2^{k_2-1}}\,\mathrm{cov}(q^2_{\boldsymbol{X},k_1}(j_1), q^2_{\boldsymbol{X},k_2}(j_2)) \\
&\leqslant \check{B}_1 + \check{B}_2,
\end{aligned}$$

where $\check{B}_1$ to be the sum over higher levels $k_2 \geqslant \log_2[(L-\sqrt{L})/(M+2)]$, while $\check{B}_2$ to be the sum over $k_2 < \log_2[(L-\sqrt{L})/(M+2)]$. By Lemma B.1 (a) (ii)

$$\begin{aligned}
\check{B}_1 &= O(1)\,\frac{1}{n}\sum_{k_1=1}^{D}\sum_{k_2=\log_2[(L-\sqrt{L})/(M+2)]}^{D}\sum_{\substack{t_1,t_2=1\\ t_1\neq t_2}}^{N}\sum_{j_1=(t_1-1)L+1}^{t_1L}\sum_{j_2=(t_2-1)L+1}^{t_1L} 2^{(4d-2)k_2} \\
&= O(n\log n)\sum_{k_2=\log_2[(L-\sqrt{L})/(M+2)]}^{D} 2^{(4d-2)k_2} = O\left(nL^{4d-2}\log n\right) = o(1)
\end{aligned}$$

by assumption.

The lower level summands $\check{B}_2 = \check{B}_{2,1} + \check{B}_{2,2} + \check{B}_{2,3} + \check{B}_{2,4}$ need to be further split similarly to the treatment of $\widetilde{A}_3$ in the proof of Theorem 3.2: $\check{B}_{2,1}$ is the sum over $|j_2 - j_1| \geqslant (M+2)2^{k_2}$ (which implies $|\lceil j_2/2^{k_2-1}\rceil 2^{k_2} - \lceil j_1^{k_1}/2^{k_1-1}\rceil| \geqslant (M+2)\,2^{k_2}$) as well as $j_s \geqslant (M-1)2^{k_s-1}$, $s = 1, 2$, (implying $\lceil j_s/2^{k_s-1}\rceil \geqslant \lceil M_{k_s}/2^{k_s}\rceil$). $\check{B}_{2,2}$ sums over $|j_2 - j_1| < (M+2)2^{k_2}$, $\check{B}_{2,3}$ sums over $j_s < (M-1)2^{k_s-1}$ as well as $j_t < n/2 - M2^{k_s-1}$ (implying $\lceil j_t/2^{k_t-1}\rceil < n/2^{k_t} - M2^{k_s-k_t}$) as well as $|j_2 - j_1| \geqslant (M+2)2^{k_2}$ with $s = 1$ and $t = 2$ as well as $s = 2$ and $t = 1$. Finally $\check{B}_{2,4}$ is the sum over $j_s < (M-1)2^{k_s-1}$ and $j_t > n/2 - M2^{k_s-1}$ with $t = 1$ and $s = 2$ as well as $t = 2$ and $s = 1$.

Using a combination of Lemma B.1 (a) (ii) and (b) (ii) and the fact that $\min(a,b) \leqslant a^\alpha b^{1-\alpha}$ for any $0 \leqslant \alpha \leqslant 1$ we obtain by an application of (B.9) and (B.10) (similarly to the treatment of $\widetilde{B}_1$ in the proof of Theorem 3.2)

$$\begin{aligned}
\check{B}_{2,1} = O(1)\,\frac{1}{n} &\sum_{k_1\leqslant k_2=1}^{\log_2[(L-\sqrt{L})/(M+2)]} \sum_{\substack{t_1,t_2=1\\ t_1\neq t_2}}^{N} \sum_{j_1=(t_1-1)L+1}^{t_1L} \sum_{j_2=(t_2-1)L+1}^{t_1L} \\
&\times \min\left(\Big||j_2-j_1| - (M+2)2^{k_2}\Big|^{4d-2}, 2^{(4d-2)k_2}\right) \\
= O(1) &\sum_{k_1\leqslant k_2}^{\log_2[(L-\sqrt{L})/(M+2)]} \frac{1}{L} \sum_{r=(M+2)2^{k_2}+1}^{2L} (r-(M+2)2^{k_2})(r-(M+2)2^{k_2})^{4d-2} \\
+ O(1) &\sum_{k_1\leqslant k_2}^{\log_2[(L-\sqrt{L})/(M+2)]} 2^{k_2} 2^{(4d-2)k_2/(2-4d)}\,\frac{1}{L} \sum_{r=(M+2)2^{k_2}+1}^{2L} (r-(M+2)2^{k_2})^{(4d-2)(1-1/(2-4d))} \\
+ O(1) &\sum_{k_1\leqslant k_2}^{\log_2[(L-\sqrt{L})/(M+2)]} \sum_{|r|\geqslant\sqrt{L}} |r|^{4d-2},
\end{aligned}$$

where we used that $L - (M+2)2^{k_2} \geqslant \sqrt{L}$. By standard calculations all three summands converge to 0 for $d < 1/4$, so that $\check{B}_{2,1} = o(1)$. Term $\check{B}_{2,2}$ can be dealt with analogously to term $\widetilde{B}_2$ in the proof of Theorem 3.2 using Lemma B.1 (a) (ii) resulting in the rate $O((\log L)/L\,\max(\log L, L^{4d}))$ on noting that for any given t_1

$$\sum_{t_2\neq t_1}\sum_{j_1,j_2} 1_{\{|j_2-j_1|<(M+2)2^{k_2}\}} = O(2^{2k_2}).$$

Terms $\check{B}_{2,3}$ can be dealt with similarly to $\widetilde{B}_3$ using Lemma B.1 c) (ii) leading to the additional rate $L\log L/n = o(1)$, while $\check{B}_{2,4}$ yield by Lemma B.1 (a) (ii) the rate $\log L/n\max(\log L, L^{4d}) = o(1)$. ■

C. Proofs of Section 4

Proof of Theorem 4.1. It holds for $J \geqslant 1$

$$
\begin{aligned}
&\mathrm{E}\, a(n,d)\,|T_n(\boldsymbol{X}) - T_{n,J}(\boldsymbol{X})| = \frac{a(n,d)}{n} \sum_{k=D-J+1}^{D} \sum_{l=1}^{n_k} \mathrm{E}\, d_k^2(l) \\
&\sim_d a(n,d) \sum_{k=D-J+1}^{D} 2^{(2d-1)k} = a(n,d) 2^{(2d-1)(D-J+1)} \frac{1-2^{(2d-1)J}}{1-2^{2d-1}} \\
&\sim_d a(n,d) n^{2d-1}\, 2^{-(2d-1)J} = 2^{(1-2d)J} \begin{cases} n^{2d-1/2}, & 0 \leqslant d < 1/4, \\ (\log n)^{-1/2}, & d = 1/4, \end{cases}
\end{aligned}
$$

where $X \sim_d Y$ means $k_d \leqslant X/Y \leqslant K_d$ for some constants $0 < k_d \leqslant K_d < \infty$ only depending on d. Because the latter term converges to zero, assertion a) is proven. For $d > 1/4$ we obtain (even for $J = 1$)

$$
\mathrm{E}\, \frac{a(n,d)}{n} d_D^2(1) \geqslant c_d n^{-2d} 2^{2dD} = c_d.
$$

■