Locally stationary long memory estimation

François Roueff, Rainer von Sachs

LOCALLY STATIONARY LONG MEMORY ESTIMATION

FRANOIS ROUEFF AND RAINER VON SACHS

Abstract. Spectral analysis of strongly dependent time series data has a long history in applications in a variety of fields, such as, e.g., telecommunication, meteorology, hydrology or, more recently, financial and economical data analysis. There exists a wide literature on parametrically or semi-parametrically modelling such processes using a long-memory parameter d, including more recent work on wavelet estimation of d. As a generalization of these latter approaches, in this work we allow the long-memory parameter d to be varying over time. Hence, we give up the somewhat restrictive assumption of second-order stationarity of the observed process (or its increments, respectively, after differencing a finite number of times). We embed our approach into the framework of locally stationary processes which, over the past decade, has been developed for weakly dependent time series with a time-varying spectral structure. In this paper we adopt a semi-parametric approach for estimating the time-varying parameter d in order to avoid fitting a parametric model, such as ARFIMA, to the observed data. We show weak consistency and a central limit theorem for our log-regression wavelet estimator of the time-dependent d in a Gaussian context. Both simulations and a real data example complete our work on providing a fairly general approach.

1. Introduction

There is a long tradition of modelling the phenomenon of long-range dependence in observed data that show a strong persistence of their correlations by long-memory processes. Such data can typically be found in the applied sciences such as hydrology, geophysics, climatology and telecommunication (e.g. teletraffic data) but recently also in economics and in finance, e.g. for modelling (realized) volatility of exchange rate data or stocks. The literature on stationary long-memory processes is huge (see e.g. the references in the recent survey paper [8]), and we concentrate here on the discussion of long-range dependence resulting from a singularity of the spectral density at zero frequency - corresponding to a slow, i.e. polynomial, decay of the autocorrelation of the data. Although a lot of (earlier) work started from a parametric approach, using e.g. the celebrated ARFIMA-like models, it occurs that since the seminal work by P. Robinson (see [16, 15]), the semi-parametric approach is known to be more robust against model misspecification: instead of using a parametric filter describing both the singularity of the spectral density at zero frequency and the ARMA-dynamics of the short memory part, only the singular behavior of the spectrum at zero is modelled by the long-memory parameter, d say, whereas the short memory part remains completely non-parametric.

Driven by the empirical observation that the correlation structure of observed (weakly or strongly dependent) data can change over time, there is a also a growing literature on modelling departures from covariance-stationarity, mainly restricted to the short-range

1991 *Mathematics Subject Classification.* Primary 62M10, 62M15, 62G05 Secondary: 60G15.

Key words and phrases. locally stationary process, long memory, semi-parametric estimation, wavelets.

dependent case. One prominent approach, that we adopt in this paper, too, is the model of local stationarity, introduced by a series of papers by R. Dahlhaus ([4, 5, 3]): in a non-parametric set-up, the spectral structure of the underlying stochastic process is allowed to be smoothly varying over time. Of course, time-varying linear processes (of ARMA type) arise as a subclass of these locally stationary processes. In order to come up with a rigorous asymptotic theory of consistency and inference, the time-dependence of the spectral density $f(u, \lambda)$ of such locally stationary processes is modelled to be in rescaled time $u \in [0, 1]$, leading to a problem of non-parametric curve estimation: increasing the sample size T of the observed time series does no more mean to look into the future but to dispose of more and more observations to identify $f(t/T, \lambda)$ locally around the "reference" rescaled time point $u \approx t/T$.

In the aforementioned spirit of semi-parametric modelling, and in contrast to the parametric approach of [2], one of the very few existing approaches on time-varying long-memory modelling, we consider in this paper a locally stationary long-range dependent process with a singularity in the spectral density at zero frequency which is parameterized by a time-varying long-memory parameter $d = d(u)$, $u \in [0, 1]$, i.e. defined in rescaled time. Our approach is a true generalization of the stationary approach in that the latter corresponds to a time-constant d for our locally stationary model. As in the case of [12], the long memory parameter is estimated by a log-regression of a series of wavelet scalograms (estimated wavelet variances per scale by summing the squared wavelet coefficients per scale over location) onto a range of scales (corresponding to the low frequency range of the spectrum). Although wavelets do not improve the estimation of d in the standard stationary context $-1/2 < d < 1/2$, their use is of interest in various practical situations (presence of trends, under and over-differenced series leading to $d \geq 1/2$ and $d \leq -1/2$ respectively), see details in [8]. However, in our work now the challenge is to *localize* the estimation of the no more constant parameter d. Wavelets are favorable in this situation since, in contrast to a Fourier analysis, they are well localized both over time and frequency, i.e. scale. The localization is achieved by smoothing over time the series of squared wavelet coefficients on each of the coarse scales which enter into the log-regression, giving raise to a *local scalogram*. We propose both a more traditional method based on two-sided kernels and also a recursive scheme of one-sided smoothing weights, adapted to the end point of the observation period.

Let us compare our approach with the few existing prepublications on time-varying long memory. A specific time-varying long memory stochastic volatility model has been treated by [10]. They use the log linear relationship of the local variance of the maximum overlap discrete wavelet transform and their scaling parameter, plus a localization with a rectangular window in coefficient domain, to estimate the time-varying long memory parameter. As mentioned already above, the work by [2] relies on a completely parametric approach for the correlation structure of the observed locally stationary time series in that the filter in the linear (although locally infinite) autoregressive representation of the process is completely determined by a finite-dimensional parameter. The very recent work by [9] finally treats two particular instances of long-memory processes with a non-constant memory parameter: seasonal processes modelled to allow singularities in the spectral density at frequencies different from zero, but using an approach with a piecewise constant long-memory parameter d over time.

Summarizing our results, the rest of the paper is organized as follows. In Section 2, we give the technical details of our locally stationary long memory model of semi-parametric type and give a series of examples of processes falling into this model. In Section 3 we define our estimators based on wavelet analysis for which we briefly recall the wavelet set-up. We define the local scalogram which is at the heart of our wavelet based estimators. We also prepare our technique of stationary approximation by defining what we call the approximating stationary *tangent process* and its wavelet spectrum, the *local wavelet spectrum*, as well as the pseudo-estimator *tangent scalogram*. We finish this section by discussing a series of smoothing weights, one- and two-sided kernels, which fulfill our given assumptions. The asymptotic properties of our proposed estimators are stated in the following Section 4. We derive a mean-square approximation of the local scalogram through the tangent scalogram (Proposition 1), followed by a control of the mean square error of the scalogram as an estimator of the local wavelet spectrum (Theorem 1) and a CLT for the tangent scalogram (Theorem 2), which finally allows us to derive a CLT for the local scalogram (Corollary 1). The results on the asymptotic behavior of the estimator of $d(u)$ are then obtained: Corollary 2 provides the rate of convergence and Theorem 3 the asymptotic normality. We pursue the paper by Section 5 on numerical examples, first simulating some ARFIMA process with a time-varying d and comparing the performance of the two-sided (triangular) kernel with the recursive weight scheme. Second, we apply the kernel estimator to a series of realized log volatilities (see also [17]), namely of the exchange rate of the YEN versus USD, from June 1986 to September 2004. We conclude in Section 6 before an appendix section presents all technical details of our derivations including our proofs.

2. Model set-up and examples

Define the difference operator $[\mathbf{\Delta} X]_k = X_k - X_{k-1}$ and $\mathbf{\Delta}^p$ recursively. This will allow $d(u)$ to take values up to $p + 1/2$ in the following model.

We adapt the approach of [4] to the case where the spectral density is allowed to have a singularity at the zero frequency. We define an array $\{X_{t,T}\}$ of Gaussian random variables, for a fixed $p = 0, 1, 2, \ldots$,

$$\mathbf{\Delta}^p X_{t,T} = \int_{-\pi}^{\pi} A^0_{t,T}(\lambda)\ \mathrm{e}^{\mathrm{i}\lambda t}\ \mathrm{d}Z(\lambda)\ , \qquad t = 1, \ldots T,\ T \geq 1\ , \tag{1}$$

where $\mathrm{d}Z(\lambda)$ are the orthonormal increments of a Wiener process Z on $[-\pi, \pi]$ and $A^0_{t,T}(\lambda)$ is an array of $L^2([-\pi, \pi])$ functions with real-valued Fourier coefficients. We further assume that there exist a function $A(u, \lambda)$ in $L^2([0, 1] \times [-\pi, \pi])$ and two constants $c > 0$ and $D < 1/2$ such that

$$\left|A^0_{t,T}(\lambda) - A(t/T, \lambda)\right| \leq c\, T^{-1}\ |\lambda|^{-D}\ , \quad 1 \leq t \leq T,\ -\pi \leq \lambda \leq \pi\ , \tag{2}$$

and

$$|A(u; \lambda) - A(v, \lambda)| \leq c\ |v - u|\ |\lambda|^{-D}\ , \quad 0 \leq u, v \leq 1,\ -\pi \leq \lambda \leq \pi\ . \tag{3}$$

These correspond to the definition of locally stationary processes introduced in [4] but with the term $|\lambda|^{-D}$ added in the upper bound to allow a singularity at the zero frequency.

This gives rise to the following time-varying *generalized* spectral density of $\{X_{t,T}\}$

$$f(u, \lambda) = |1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{-2p}\ \ |A(u; \lambda)|^2\ . \tag{4}$$

Definition 1. *We say that the process* $\{X_{t,T},\ t = 1, \ldots, T,\ T \geq 1\}$ *has* local memory parameter $d(u) \in (-\infty, p + 1/2)$ *at time* $u \in [0, 1]$ *if it satisfies (1), (2) and (3) and its* *generalized spectral density $f(u,\lambda)$ defined by (4) satisfies the following semi-parametric type condition:*

$$f(u,\lambda) = |1-\mathrm{e}^{-\mathrm{i}\lambda}|^{-2d(u)}\ f^*(u,\lambda)\ , \tag{5}$$

with $f^*(u,0) > 0$ and

$$|f^*(u,\lambda) - f^*(u,0)| \leq C\ f^*(u,0)\ |\lambda|^\beta,\ \lambda \in [-\pi,\pi]\ , \tag{6}$$

where $C > 0$ and $\beta \in (0,2]$.

The following intuitive definition will be also useful when developing our estimation theory using stationary approximations. For any $u \in [0,1]$ one may define a *tangent* stationary process for the p-th increment

$$\mathbf{\Delta}^p X_t(u) = \int_{-\pi}^{\pi} A(u;\lambda)\ \mathrm{e}^{\mathrm{i}\lambda t}\ \mathrm{d}Z(\lambda)\ , \tag{7}$$

whose spectral density is $|1-\mathrm{e}^{-\mathrm{i}\lambda}|^{2p} f(u,\lambda)$. Equivalently, we may write

$$\mathbf{\Delta}^p X_t(u) = \sum_{k\in\mathbb{Z}} a_k(u)\ \varepsilon_{t-k}\ , \tag{8}$$

where $\{\varepsilon_t\}$ are defined by

$$\varepsilon_t = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\lambda t}\ \mathrm{d}Z(\lambda)\ ,$$

(and thus are i.i.d. $\sim\ \mathcal{N}(0,1)$) and $\{a_k(u),\ k\in\mathbb{Z}\}$ is the ℓ^2 sequence defined by

$$A(u;\lambda) = \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} a_k(u)\ \exp(-\mathrm{i}\lambda\ k)\ .$$

A special case is to assume $A^0_{t,T}(\lambda) = A(t/T,\lambda)$ so that

$$\mathbf{\Delta}^p X_{t,T} = \int_{-\pi}^{\pi} A(t/T;\lambda)\ \mathrm{e}^{\mathrm{i}\lambda t}\ \mathrm{d}Z(\lambda)\ , \tag{9}$$

In this case $\{X_{t,T}\}$ has locally stationary p-order increments with the following MA(∞)-representation

$$\mathbf{\Delta}^p X_{t,T} = \sum_{k\in\mathbb{Z}} a_k(t/T)\ \varepsilon_{t-k}\ . \tag{10}$$

We now give a small series of examples, adapted from [12] to the time varying setting.

Example 1 (tvFBM(H)). The *Fractional Brownian motion* (FBM) process $\{B_H(k)\}_{k\in\mathbb{Z}}$ with Hurst index $H \in (0,1)$ is a discrete-time version of $\{B_H(t), t\in\mathbb{R}\}$, a Gaussian process with mean zero and covariance

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}\left\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right\}\ .$$

The spectral density of $\{\mathbf{\Delta} B_H(k)\}_{k\in\mathbb{Z}}$ is then given by $\lambda \mapsto |1-\mathrm{e}^{-\mathrm{i}\lambda}|^{-2H+1} f_{\mathrm{FBM}}(\lambda;H)$, where

$$f_{\mathrm{FBM}}(\lambda;H) = \left|\frac{2\sin(\lambda/2)}{\lambda}\right|^{2H+1} + |2\sin(\lambda/2)|^{2H+1} \sum_{k\neq 0} |\lambda + 2k\pi|^{-2H-1}\ . \tag{11}$$

The *time varying Fractional Brownian motion* (tvFBM) is defined by (9), (4) and (5) with $p=1$, $d(u) = H(u)+1/2 \in (1/2,3/2)$ and $f^*(u,\lambda) = f_{\mathrm{FBM}}(\lambda;H(u))$, where H is a Lipschitz mapping of $[0,1]$ into a subset of $(0,1)$. Then (6) holds with $\beta = (2H(u)+1) \wedge 2$. The corresponding non-negative local transfer function is

$$A(u,\lambda) = |1-\mathrm{e}^{-\mathrm{i}\lambda}|^{1/2-H(u)} \sqrt{f_{\mathrm{FBM}}(\lambda; H(u))} \ .$$

In this case, by Lemma 3 in the appendix, (3) holds for any $D > \sup_u H(u) - 1/2$.

Example 2 (tvFGN(H)). The *time varying fractional Gaussian noise* (tvFGN) is defined similarly as the tvFBM by $f^*(u,\lambda) = f_{\mathrm{FBM}}(\lambda; H(u))$ but with $p=0$ and $d(u) = H(u) - 1/2 \in (-1/2, 1/2)$.

Example 3 (tvARFIMA($0,d,0$)). If $f^*(u;\lambda) = 1$, then, for a given differentiation order $p \geq 0$, one may set

$$A(u;\lambda) = (1-\mathrm{e}^{-\mathrm{i}\lambda})^{-d(u)+p} \ ,$$

with d a Lipschitz function from $[0,1]$ to $(-\infty, p+1/2)$. Lemma 3 yields (3) for any $D > \sup_u d(u) - p$. In this case the tangent process (8) is a *fractionally integrated noise* (ARFIMA($0, d(u)-p, 0$)) process, hence we call X a *time varying fractionally integrated noise* (tvARFIMA($0,d,0$)) process.

Example 4 (tvARFIMA(q,d,r)). The *time varying autoregressive fractionally integrated moving average* (tvARFIMA(q,d,r)) process is similar to the tvARFIMA($0,d,0$) but we replace $f^*(u;\lambda) = 1$ by

$$f^*(u;\lambda) = \frac{\sigma^2(u)}{2\pi} \frac{1+\sum_{k=1}^r \theta_k(u)\mathrm{e}^{-\mathrm{i}\lambda}}{1-\sum_{k=1}^r \phi_k(u)\mathrm{e}^{-\mathrm{i}\lambda}} \ , \tag{12}$$

where $\sigma : [0,1] \to \mathbb{R}_+$, $\phi = [\phi_1 \ \ldots \ \phi_q]^T : [0,1] \to \mathbb{R}^q$ and $\theta = [\theta_1 \ \ldots \ \theta_r]^T : [0,1] \to \mathbb{R}^r$ are Lipschitz functions. The conditions (6) and (3) continue to hold with the same $\beta = 2$ and the same $D > \sup_u d(u) - p$, respectively.

In order to verify Condition (2) trivially, the simplest definition of $\{\boldsymbol{\Delta}^p X_{t,T}\}$ in all the previous examples is to take $A^0_{t,T}(\lambda) = A(t/T, \lambda)$, that is to set (9), as will be done for our simulated tvARFIMA in Section 5. However, one might also want to use a different transfer function $A^0_{t,T}$ in (1), provided that Condition (2) holds. Such approximated tvMA(∞) representation is motivated by the tvAR(p) process, which satisfies the recursion

$$X_{t,T} - \sum_{k=1}^{p} \phi_k(t/T) X_{t-k,T} = \sigma(t/T)\ \epsilon_t, \quad 1 \leq t \leq T \ ,$$

along with appropriate initial conditions. It has been shown in [4] that such non-stationary process does not satisfy a representation of the form (9) (with $p=0$) but it does satisfy (1) and (2) (with $p = D = 0$).

3. Estimation method based on wavelet analysis

3.1. **Discrete wavelet transform (DWT).**

Following the approach presented in [12] for the estimation of the memory parameter of a stationary sequence, we compute the discrete wavelet transform (DWT) of $\{X_{t,T},\ 1 \leq t \leq T\}$ (in discrete time) for a given scale function ϕ and wavelet ψ. We denote by $\{W_{j,k;T}\,;\, j \geq 0, k \in \mathbb{Z}\}$ the wavelet coefficients of the process $\{X_{t,T},\ 1 \leq t \leq T\}$,

$$W_{j,k;T} = \sum_{t=1}^{T} h_{j,2^j k - t} X_{t,T}, \quad k = 0, \ldots, T_j - 1 \ , \tag{13}$$

where $\{h_{j,t}, \ , t \in \mathbb{Z}\}$ denotes the wavelet detail filter at scale j associated to ϕ and ψ through the relation

$$h_{j,t} = 2^{-j/2} \int_{-\infty}^{\infty} \phi(u+t)\psi(2^{-j}u) \, \mathrm{d}u \ ,$$

and T_j the number of available wavelet coefficients at scale j, which satisfies

$$T2^{-j} - c \leq T_j \leq T2^{-j}, \quad \text{for some constant } c \text{ independent of } j \geq 0 \ . \tag{14}$$

The filter $h_{j,\cdot}$ and T_j are defined so that the support $\{t \ : \ h_{j,2^jk-t} \neq 0\}$ is included in $\{1, \ldots T\}$ for $k = 0, \ldots, T_j - 1$. Observe that here j denotes the scale index of the wavelet coefficient and k its position index. We use the convention that a large j corresponds to a coarse scale. Let us define

$$H_j(\lambda) = \sum_{t \in \mathbb{Z}} h_{j,t} \mathrm{e}^{-\mathrm{i}t\lambda} \tag{15}$$

the corresponding filter transfer function. The following conditions on the wavelet ψ and scale function ϕ are assumed to hold for a positive integer M and a real α.

(W-1) ϕ and ψ are compactly-supported, integrable, $\int_{-\infty}^{\infty} \phi(t) \, \mathrm{d}t = 1$ and $\int_{-\infty}^{\infty} \psi^2(t) \, \mathrm{d}t = 1$.

(W-2) There exists $\alpha > 1$ such that $\sup_{\xi \in \mathbb{R}} |\hat{\psi}(\xi)| \, (1+|\xi|)^{\alpha} < \infty$, where $\hat{\psi}(\xi) = \int_{-\infty}^{\infty} \psi(t) \, \mathrm{e}^{-\mathrm{i}\xi t} \mathrm{d}t$ denotes the Fourier transform of ψ.

(W-3) The function ψ has M vanishing moments, $\int_{-\infty}^{\infty} t^m \psi(t) \, \mathrm{d}t = 0$ for all $m = 0, \ldots, M - 1$

(W-4) The function $\sum_{k \in \mathbb{Z}} k^m \phi(\cdot - k)$ is a polynomial of degree m for all $m = 0, \ldots, M-1$.

Under (W-3) and (W-4), the filter can be interpreted as the convolution of the $\mathbf{\Delta}^M$ filter with a finite impulse response filter. Hence if $M \geq p$, Equation (13) may be written as

$$W_{j,k;T} = \sum_{t=1}^{T} \tilde{h}_{j,2^jk-t} (\mathbf{\Delta}^p X)_{t,T}, \quad k = 0, \ldots, T_j - 1 \ ,$$

where $h_{j,\cdot} = \tilde{h}_{j,\cdot} * \mathbf{\Delta}^p$. In particular, we have

$$\tilde{H}_j(\lambda) = \sum_{t \in \mathbb{Z}} \tilde{h}_{j,t} \mathrm{e}^{-\mathrm{i}t\lambda} = H_j(\lambda)(1 - \mathrm{e}^{\mathrm{i}\lambda})^{-p} \ . \tag{16}$$

3.2. **Local wavelet spectrum, local scalogram, tangent scalogram, and final estimator.** Recall that $f(u, \cdot)$ in (4) can be interpreted as a *local generalized spectral density* at rescaled time $u \in [0,1]$. Hence, as in the stationary setting used in [12], for each such u, we may define a *local wavelet spectrum* $\sigma^2(u) = \{\sigma_j^2(u), \, j \geq 0\}$, where for each $j \geq 0$, $\sigma_j^2(u)$ is the variance of the wavelet coefficients at scale index j of a process with generalized spectral density $f(u, \cdot)$. This variance is well defined under the assumption $M \geq p$ because in this case the wavelet coefficients at given scale are weakly stationary. Moreover, by (4) and (16),

$$\sigma_j^2(u) = \int_{-\pi}^{\pi} |H_j(\lambda)|^2 \ f(u;\lambda) \, \mathrm{d}\lambda = \int_{-\pi}^{\pi} \left| \tilde{H}_j(\lambda) \ A(u;\lambda) \right|^2 \, \mathrm{d}\lambda \ .$$

One way to compute the wavelet spectrum is to use the tangent process $\mathbf{\Delta}^p X_t(u)$ defined in (7). We first define the wavelet coefficients of the tangent stationary process at any

$u \in [0,1]$, namely,

$$W_{j,k}(u) = \sum_{t=1}^{T} \tilde{h}_{j,2^j k-t} (\mathbf{\Delta}^p X)_t(u) \tag{17}$$

$$= \int_{-\pi}^{\pi} \tilde{H}_j(\lambda)\ A(u;\lambda)\ \mathrm{e}^{\mathrm{i}\lambda 2^j k}\ \mathrm{d}Z(\lambda)\ , \quad k = 0, \dots, T_j - 1\ . \tag{18}$$

these wavelet coefficients are indeed those of a process with generalized spectral density $f(u,\cdot)$. Thus the above definition gives

$$\sigma_j^2(u) = \mathbb{E}\left[W_{j,k}^2(u)\right]\ . \tag{19}$$

An important tool for the estimation of the long memory is the *scalogram* (first introduced in this context by [18] and developed by [1]) defined as

$$\widehat{\sigma}_j^2 = T_j^{-1} \sum_{k=0}^{T_j - 1} W_{j,k}^2\ .$$

Here to cope with local stationarity, we will need a *local scalogram* for estimating the local wavelet spectrum, namely, for a given $u \in [0,1]$,

$$\widehat{\sigma}_{j,T}^2(u) = \sum_{k=0}^{T_j-1} \gamma_{j,T}(k) W_{j,k;T}^2\ , \tag{20}$$

where $\{\gamma_{j,T}(k)\}$ are some non-negative weights localized at indices $k \approx uT_j$ and normalized in such a way that

$$\sum_{k=0}^{T_j-1} \gamma_{j,T}(k) = 1\ . \tag{21}$$

The localization property will be expressed by imposing a bound on the increase rate of the following quantity (see equation (30))

$$\Gamma_q(u;j,T) = \sum_{k=0}^{T_j-1} |\gamma_{j,T}(k)|\ \ |k - Tu2^{-j}|^q\ , \tag{22}$$

as $T \to \infty$ for appropriate values of the exponent q.

An important tool for studying the local scalogram is the *tangent scalogram* defined as

$$\widetilde{\sigma}_{j,T}^2(u) = \sum_{k=0}^{T_j-1} \gamma_{j,T}(k) W_{j,k}^2(u)\ . \tag{23}$$

We note that this definition is similar to that of the local scalogram in (20) but with the wavelet coefficients $W_{j,k;T}$ replaced by the tangent wavelet coefficients $W_{j,k}^2(u)$ defined in (17). The tangent scalogram is not an estimator since it cannot be computed from the observations $X_{1,T}, \dots, X_{T,T}$. However, it provides useful approximations of the local scalogram.

We conclude this section by deriving an *estimator of the time-varying long memory parameter*. The local wavelet spectrum is related to the local memory parameter $d(u)$ by the asymptotic property $\sigma_j^2(u) \sim c2^{2d(u)j}$ as $j \to \infty$. This property will be made more precise when we study the bias, see the relation (37) below. An estimator of $d(u)$ is obtained by a linear regression of $(\log \widehat{\sigma}^2_{j,T}(u))_{j=L,\dots,L+\ell}$ with respect to $j = L, \dots, L+\ell$, where ℓ is the number of scales used in the regression and L is the lowest scale index used in the regression. Let $\mathbf{w}$ be a vector $\mathbf{w} = [w_0, \dots, w_\ell]^T$ satisfying

$$\sum_{i=0}^{\ell} w_i = 0 \quad \text{and} \quad 2\log(2)\sum_{i=0}^{\ell} i w_i = 1 \; . \tag{24}$$

The local estimator of $d(u)$ is defined as

$$\hat{d}_T(L) = \sum_{j=L}^{L+\ell} w_{j-L} \log\left(\widehat{\sigma}^2_{j,T}(u)\right) \; . \tag{25}$$

3.3. **Conditions on the weights $\gamma_{j,T}(k)$ and examples.** Let us now precise our conditions on the weights $\gamma_{j,T}(k)$. Denote, for any $0 \le i \le j$, $v \in \{0, \dots, 2^i - 1\}$ and $\lambda \in \mathbb{R}$,

$$\Phi_{j,T}(\lambda; i, v) = \sum_{l \in \mathcal{T}_j(i,v)} \gamma_{j-i,T}(2^i l + v) \mathrm{e}^{\mathrm{i} l \lambda} \; , \tag{26}$$

where

$$\mathcal{T}_j(i,v) = \left\{ l \,:\, 0 \le l < 2^{-i}(T_{j-i} - v) \right\} \; . \tag{27}$$

We moreover define

$$\delta_{j,T} = \sup_{k=0,\dots,T_j-1} |\gamma_{j,T}(k)| \; . \tag{28}$$

The weights $\gamma_{j,T}(k)$ must satisfy an appropriate asymptotic behavior as $T \to \infty$ for obtaining a consistent estimator of $d(u)$. More precisely, the following assumption will be required.

Assumption 1. The index j depends on T so that the weights $(\gamma_{j,T}(k))_k$ satisfy the following asymptotic properties as $T \to \infty$.

(i) We have $\delta_{j,T} \to 0$, and for any fixed integer i, $\delta_{j+i,T} \sim 2^i \delta_{j,T}$.

(ii) For all $i, i' \ge 0$, $v \in \{0, \dots, 2^i - 1\}$ and $v' \in \{0, \dots, 2^{i'} - 1\}$, there exists a constant $V = V(i, v; i', v')$ such that

$$\delta_{j,T}^{-1} \int_{-\pi}^{\pi} \Phi_{j,T}(\lambda; i, v) \overline{\Phi_{j,T}(\lambda; i', v')} \, \mathrm{d}\lambda \to V(i, v; i', v') \; . \tag{29}$$

(iii) For all $\eta > 0$, $i \ge 0$ and $v \in \{0, \dots, 2^i - 1\}$, we have

$$\delta_{j,T}^{-1/2} \sup_{\eta \le |\lambda| \le \pi} |\Phi_{j,T}(\lambda; i, v)| \to 0 \; .$$

(iv) For $q = 0, 1, 2$, we have

$$\Gamma_q(u; j, T) = O\left((\delta_{j,T})^{-q}\right) \; , \tag{30}$$

where $\Gamma_q(u; j, T)$ is defined in (22).

We provide several examples of weights satisfying this assumption below. In these examples, the weights $\gamma_{j,T}(k)$, $k = 0, \dots, T_j$, are entirely determined by T_j and a bandwidth parameter b_T and

$$\delta_{j,T}^{-1} \asymp \mathrm{b}_T T_j \sim \mathrm{b}_T T 2^{-j} \; . \tag{31}$$

In kernel estimation, one may interpret the bandwidth parameter b_T as the proportion of wavelet coefficients used for the estimation of the local scalogram $\widehat{\sigma}^2_{j,T}(u)$ at given scale j

and position u, among the T_j wavelet coefficients available at scale j from T observations $X_{1,T}, \ldots, X_{T,T}$. Lemmas 4 and 5 show that, for these examples, Assumption 1 is satisfied as soon as $T_j \to \infty$ and $\mathrm{b}_T T_j \to 0$, except in the non-compactly supported case (K-3) in Lemma 4 where we assume in addition that $T_j \exp(-c' \mathrm{b}_T T_j) = O(1)$ for any $c' > 0$, which holds in the typical asymptotic setting $\mathrm{b}_T \asymp T_j^{-\zeta}$ with $\zeta \in (0,1)$.

Example 5 (Two-sided kernel weights). For $u \in (0,1)$, one has a number of observations before rescaled time u and after rescaled time u both tending to infinity. Thus we may use a two-sided kernel to localize the memory parameter estimator around u. For a given bandwidth $b = b_T$, the corresponding weights are given by

$$\gamma_{j,T}(k) = \rho_{j,T}^{-1} \, K((uT_j - k)/(\mathrm{b}_T T_j)) \; , \tag{32}$$

where K is a non-negative symmetric function and $\rho_{j,T}$ is a normalizing term so that (21) holds. In the last display we see that b_T is the bandwidth in a rescaled time sense while $\mathrm{b}_T T_j$ is the corresponding bandwidth in the sense of location indices $k = 0, 1, 2, \ldots, T_j$ at scale j . Lemma 4 in the appendix shows that Assumption 1 holds for a wide variety of choices for K.

Example 6 (Recursive weights). By *recursive weights*, we here mean that, given T, L and $\mathbf{w}$, the possibility of computing $\widehat{\sigma}^2_{j,T}(u)$ by successive simple linear processing involving a finite number of operations after each new observations $X_{t;T}$ as t grows from $t = 1$ to $t = T$.

Because the DWT is defined as a succession of finite filtering and decimation, it is indeed possible to compute $W_{j,k;T}$ online for all $j \in \{L, \ldots, L + \ell\}$ and $k \in \{0, \ldots, T_j\}$. Then an online implementation of the local *recursive* scalogram can be done by setting

$$\widehat{\sigma}^2_{j,-1;T} = 0 \; , \quad j \in \{L, \ldots, L + \ell\},$$

and, using the following recursive equation for all $j \in \{L, \ldots, L+\ell\}$ and $t \in \{0, \ldots, T_j - 1\}$,

$$\widehat{\sigma}^2_{j,t;T} = \exp(-(\mathrm{b}_T T_j)^{-1}) \, \widehat{\sigma}^2_{j,t-1;T} + W^2_{j,t;T} \; ,$$

where $(\mathrm{b}_T T_j)^{-1}$ is the exponential *forgetting exponent* corresponding to the *bandwidth parameter* b_T. For any $u \in (0,1]$, we define a local recursive scalogram by

$$\widehat{\sigma}^2_{j;T}(u) = \rho_{j,T}^{-1} \widehat{\sigma}^2_{j,[uT_j]-1;T} \; ,$$

where $[a]$ denotes the integer part of a and

$$\rho_{j,T} = \sum_{k=0}^{[uT_j]-1} \mathrm{e}^{-k/(\mathrm{b}_T T_j)} = \frac{1 - \mathrm{e}^{-[uT_j]/(\mathrm{b}_T T_j)}}{1 - \mathrm{e}^{-(\mathrm{b}_T T_j)^{-1}}} \; . \tag{33}$$

Hence (20) and (21) hold with

$$\gamma_{j,T}(k) = \rho_{j,T}^{-1} \mathrm{e}^{-([uT_j]-1-k)/(\mathrm{b}_T T_j)} \mathbb{1}_{[0,uT_j-1)}(k) \; . \tag{34}$$

Observe that these weights are one-sided by construction, since only the observations before rescaled time u are used for estimating $d(u)$. This is the reason why we may take $u \in (0,1]$. Lemma 5 shows that Assumption 1 holds for these weights, provided that $\mathrm{b}_T \to 0$ and $T_j \mathrm{b}_T \to \infty$.

4. Asymptotic properties

We study the asymptotic properties of $\widehat{d}_T(L)$ defined by (25) as $L, T \to \infty$ in such a way that Assumption 1 holds for each $j = L, L+1, \ldots, L+\ell$ and for the chosen weights $\gamma_{j,T}(k)$. We provide further conditions on $L, T, \delta_{L,T}$ under which consistency holds and derive the corresponding rate of convergence (Corollary 2). Under strengthened conditions, we further show that $\widehat{d}_T(L)$ is asymptotically normal (Theorem 3). These results essentially follow from asymptotic results on the tangent scalogram (Theorem 2, Relations (37) and (58)) and approximation results on the local scalogram (Proposition 1) based on the tangent scalogram.

4.1. Asymptotic properties of the local scalogram.

In order to derive asymptotic results for $\widehat{\sigma}^2_{j,T}(u)$, we first establish a bound on the error made when approximating $\widehat{\sigma}^2_{j,T}(u)$ by $\widehat{\sigma}^2_{j,T}(u)$.

Proposition 1. *Let $u \in [0,1]$. Assume (W-1)–(W-4) hold with $M \geq p \vee (d(u) - 1/2)$ and $\alpha > 1/2 - d(u)$. Suppose moreover that Assumption 1 (iv) holds. Then, the following approximation holds.*

$$\mathbb{E}\left[\left(\widehat{\sigma}^2_{j,T}(u) - \widetilde{\sigma}^2_{j,T}(u)\right)^2\right] = O\left(2^{(6+4p)j} T^{-4} \delta^{-4}_{j,T} + 2^{(3+2p+2d(u))j} T^{-2} \delta^{-2}_{j,T}\right) . \tag{35}$$

Next, we derive a bound of the mean square error for estimating $f^*(u,0)\kappa(d(u))\ 2^{2jd(u)}$ using the estimator $\widehat{\sigma}^2_{j,T}(u)$, where κ is the function defined by

$$\kappa(d) = \int_{-\infty}^{\infty} |\xi|^{-2d}\, |\hat{\psi}(\xi)|^2 \,\mathrm{d}\xi\ , \quad 1/2 - \alpha < d < M + 1/2\ . \tag{36}$$

In fact as the estimator $\hat{d}_T(L)$ is defined in (25) using $\widehat{\sigma}^2_{j,T}(u)$ with $j = L + i$, $i = 0, \ldots, \ell$, and as $L, T \to \infty$, it will be convenient to normalize these quantities by $2^{2Ld(u)}$, so that $f^*(u,0)\kappa(d(u))\ 2^{2jd(u)}/2^{2Ld(u)} = f^*(u,0)\kappa(d(u))\ 2^{2id(u)}$ does not depend on L.

Theorem 1. *Let $u \in [0,1]$. Assume (W-1)–(W-4) hold with $M \geq p \vee d(u)$ and $\alpha > (1+\beta)/2 - d(u)$. Then we have, as $j \to \infty$,*

$$\sigma^2_j(u) = f^*(u,0)\kappa(d(u))\ 2^{2jd(u)} \left\{1 + O\left(2^{-\beta j}\right)\right\}\ . \tag{37}$$

Suppose moreover that Assumption 1 holds and that

$$2^{(3+2\{p-d(u)\})L} T^{-2} \delta^{-2}_{L,T} \to 0\ . \tag{38}$$

Then we have for $j = L + i$ with $i = 0, \ldots, \ell$,

$$\mathbb{E}\left[(2^{-2Ld(u)}\widehat{\sigma}^2_{j,T}(u) - f^*(u,0)\kappa(d(u))\ 2^{2id(u)})^2\right] = O\left(\delta_{L,T} + 2^{(3+2\{p-d(u)\})L} T^{-2} \delta^{-2}_{L,T} + 2^{-2\beta L}\right) \tag{39}$$

Using the approximation result stated in Proposition 1, we may also wish to obtain a central limit theorem (CLT) for the local scalogram. To this end, we must first derive a CLT for the tangent scalogram. Define, for any integer $\ell \geq 0$ and $d \in (1/2 - \alpha, M]$ the

2^ℓ-dimensional cross spectral density $\mathbf{D}_{\infty,\ell}(\lambda;d) = [\mathbf{D}_{\infty,\ell,v}(d)]_{v=0,\ldots,2^\ell-1}$ of the DWT of the generalized fractional Brownian motion (see [12]) by

$$\mathbf{D}_{\infty,\ell}(\lambda;d) = \sum_{l\in\mathbb{Z}} |\lambda+2l\pi|^{-2d}\, \mathbf{e}_\ell(\lambda+2l\pi)\, \overline{\hat\psi(\lambda+2l\pi)}\hat\psi(2^{-\ell}(\lambda+2l\pi)) ,$$

where for all $\xi \in \mathbb{R}$,

$$\mathbf{e}_\ell(\xi) = 2^{-\ell/2}\,[1, \mathrm{e}^{-\mathrm{i}2^{-\ell}\xi}, \ldots, \mathrm{e}^{-\mathrm{i}(2^\ell-1)2^{-\ell}\xi}]^T .$$

In other words $\mathbf{D}_{\infty,\ell}(\lambda;d)$ is a vector with entries

$$\mathbf{D}_{\infty,\ell,v}(\lambda;d) = 2^{-\ell/2} \sum_{l\in\mathbb{Z}} |\lambda+2l\pi|^{-2d}\, \mathrm{e}^{-\mathrm{i}\,v\,2^{-\ell}(\lambda+2l\pi)}\, \overline{\hat\psi(\lambda+2l\pi)}\hat\psi(2^{-\ell}(\lambda+2l\pi)), \quad 0\le v < 2^\ell .$$

We can now state the CLT for the tangent scalogram.

Theorem 2. *Let $u\in[0,1]$. Suppose that (W-1)–(W-4) hold with $M \ge p\vee d(u)$, $\alpha > 1/2-d(u)$ and that Assumption 1 (i)–(iii) holds. Then, for any $\ell\ge 0$, the following weak convergence holds.*

$$\left(\widetilde S_L(u) - \mathbb{E}\left[\widetilde S_L(u)\right]\right) \Rightarrow \mathcal{N}(0, (f^*(u,0))^2\Sigma(u)) , \tag{40}$$

where

$$\widetilde S_L(u) = 2^{-2Ld(u)}\delta_{L,T}^{-1/2}[\widetilde\sigma^2_{L,T}(u)\ \widetilde\sigma^2_{L+1,T}(u)\ \ldots\ \widetilde\sigma^2_{L+\ell,T}(u)]^T . \tag{41}$$

and $\Sigma(u)$ is the $(\ell+1)\times(\ell+1)$ symmetric matrix defined by

$$\Sigma_{i,i'}(u) = 2\ 2^{\{1+4d(u)\}i} \sum_{v=0}^{2^{i-i'}-1} V(i-i',v) \int_{-\pi}^{\pi} \left|\mathbf{D}_{\infty,i-i',v}(\lambda;d(u))\right|^2 \mathrm{d}\lambda, \quad 0\le i'\le i\le \ell , \tag{42}$$

with $V(i-i',v) = V(0,0;i-i',v)$ defined in (29).

Applying Proposition 1 and Theorem 2, we immediately get the following result.

Corollary 1. *Let $u\in[0,1]$. Assume (W-1)–(W-4) hold with $M\ge p\vee d(u)$, $\alpha > 1/2-d(u)$. Suppose moreover that Assumption 1 holds and*

$$2^{(3+2\{p-d(u)\})L}T^{-2}\delta_{L,T}^{-3} \to 0 . \tag{43}$$

Then, for any $\ell\ge0$, the following weak convergence holds.

$$\left(\widehat S_L(u) - \mathbb{E}\left[\widehat S_L(u)\right]\right) \Rightarrow \mathcal{N}(0, (f^*(u,0))^2\Sigma(u)) , \tag{44}$$

where

$$\widehat S_L(u) = 2^{-2Ld(u)}\delta_{L,T}^{-1/2}[\widehat\sigma^2_{L,T}(u)\ \widehat\sigma^2_{L+1,T}(u)\ \ldots\ \widehat\sigma^2_{L+\ell,T}(u)]^T . \tag{45}$$

and $\Sigma(u)$ is the $(\ell+1)\times(\ell+1)$ symmetric matrix defined by (42).

4.2. **Asymptotic properties of the estimator $\widehat{d}_T(L)$.** The following result establishes the consistency of the estimator $\widehat{d}_T(L)$ defined in (25) with $\mathbf{w} = [w_0, \dots, w_\ell]^T$ fulfilling (24) and provides a rate of convergence.

Corollary 2. *Under the same assumptions as Theorem 1, if moreover $L \to \infty$, then we have*

$$\widehat{d}_T(L) = d(u) + O_p\left(\delta_{L,T}^{1/2} + 2^{(3/2+\{p-d(u)\})L} T^{-1} \delta_{L,T}^{-1} + 2^{-\beta L}\right) = d(u) + o_p(1) . \qquad (46)$$

Let us determine the optimal rate of convergence of $\widehat{d}_T(L)$ towards $d(u)$. By balancing the three terms in the right-hand side of (46), we find that for $2^L \asymp T^{2/(3+6\beta-2d(u)+2p)}$ and $\mathrm{b}_T \asymp (T_L \delta_{L,T})^{-1} \asymp T^{(2d(u)-2p-2\beta-1)/(3+6\beta-2d(u)+2p)}$, these three terms are asymptotically of the same order. Hence for this choice of the lowest scale L and the bandwidth b_T (recall that $\delta_{L,T}^{-1} \asymp \mathrm{b}_T T 2^{-L} \to \infty$), we get

$$\widehat{d}_T(L) = d(u) + O_p\left(T^{-2\beta/(3+6\beta+2\{p-d(u)\})}\right) .$$

We observe that the rate of convergence depends on the unknown parameter $d(u)$. The dependence in $d(u)$ comes from the approximation result (35), which appears in (46) in the term $2^{(3/2+\{p-d(u)\})L} T^{-1} \delta_{L,T}^{-1}$. Other error terms in (46) have rates not depending on $d(u)$, which is consistent with the facts that 1) the rate of convergence does not depend on d in the stationary case [12, Theorem 2], and 2) these two terms come from the tangent stationary approximation. On the other hand, the term $2^{(3/2+\{p-d(u)\})L} T^{-1} \delta_{L,T}^{-1}$ may seem unusual for estimating the time-varying parameter for local-stationary processes. For instance, for a time-varying AR (tvAR) process with a Lipschitz-continuous parameter corresponding to (2) with $D = 0$, the approximation error due to non-stationarity yields the error term $\mathrm{b}_T \asymp (T\delta_{L,T})^{-1}$. Indeed this corresponds to the term $(n\mu)^{-\beta}$ with $\beta = 1$ in [11, Theorem 2] which is shown to yield a rate optimal convergence in Theorem 4 of the same reference. Our error term is always larger as it includes the additional multiplicative term $2^{(3/2+\{p-d(u)\})L}$ which tends to ∞ since $p - d(u) > -1/2$ and $L \to \infty$. Although we cannot assert that our rate is optimal, it can be explained as follows. In contrast to the tvAR process, our setting is locally semi-parametric, which implies to let L tend to ∞ in order to capture the low frequency behavior driven by the memory parameter d. It is thus reassuring that if L were allowed to remain fix our error bound would be of the same order as for the locally parametric setting. The fact that letting $L \to \infty$ decreases the rate of convergence is not surprising as the low frequency behavior implies large lags in the process, which naturally worsens the quality of the local stationary approximations. To conclude this discussion, it is interesting to note that the wavelet estimation of the memory parameter of a non-Gaussian process may also yield a rate of convergence depending on the unknown parameter. It is indeed the case for the infinite-source Poisson process, see [7, Remark 4.2].

We now state the asymptotic normality of the estimator, which mainly follows by applying Proposition 1, Theorem 2, the bound (37) and the δ–method as in [13].

Theorem 3. *Let $u \in [0, 1]$. Assume (W-1)–(W-4) hold with $M \geq p \vee d(u)$, $\alpha > (1+\beta)/2 - d(u)$. Suppose moreover that Assumption 1 holds and*

$$2^{(3+2\{p-d(u)\})L} T^{-2} \delta_{L,T}^{-3} \to 0 \quad \text{and} \quad 2^{-2\beta L} \delta_{L,T}^{-1} \to 0 . \qquad (47)$$

Then, the following weak convergence holds:

$$\delta_{L,T}^{-1/2}(\widehat{d}_T(L) - d(u)) \Rightarrow \mathcal{N}(0, \mathcal{V}(u)) \ , \tag{48}$$

where $\widehat{d}_T(L)$ *is defined by (25) and*

$$\mathcal{V}(u) = \frac{1}{\kappa^2(d(u))} \sum_{i,i'=0}^{\ell} \Sigma_{i,i'}(u) 2^{-2(i+i')d(u)} w_i w_{i'} \ .$$

with $\Sigma(u)$ *and* $\kappa(d(u))$ *defined by (42) and (36), respectively.*

5. Numerical examples

We used a Daubechies wavelet with $M = 2$ vanishing moments and Fourier decay $\alpha = 1.34$ (see [8]). Hence our asymptotic results hold for $-0.84 < (1+\beta)/2 - \alpha < d(u) \le M = 2$ (the left bound 0.84 corresponds to choose β arbitrarily small). In particular $d(u)$ will be allowed to take values beyond the unit root case ($d(u) \ge 1$).

5.1. **Simulated data.** We simulate a $T = 2^{12}$-long sample $X_{1,T}, \ldots, X_{T,T}$ of a tvARFIMA(1,d,0) process which has a local spectral density given by (12) with $\sigma \equiv 1$, $\phi_1 \equiv 0.8$ and

$$d(u) = (1 - \cos(\pi u/2))/3, \quad u \in [0,1] \ .$$

The obtained simulated data is represented in Figure 1. We compute the local estimator

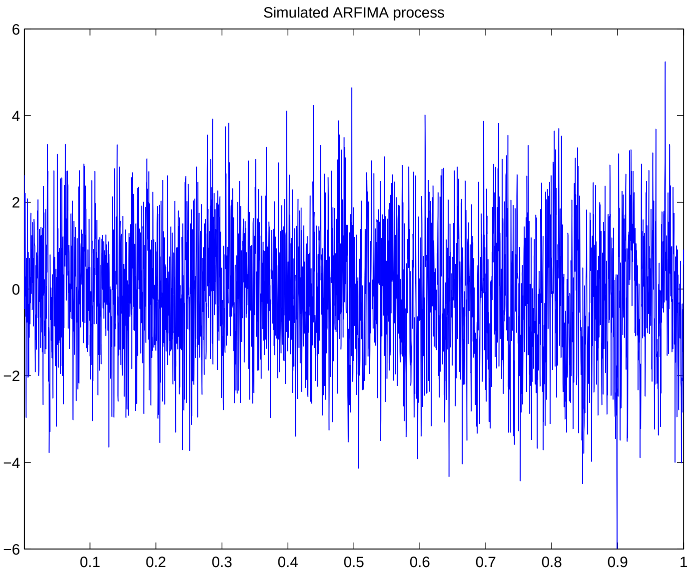

Figure 1. A simulated tvARFIMA(1,d,0) of length $T = 2^{12}$.

$\widehat{\sigma}^2_{j,T}(u)$ defined in (20) with $\{\gamma_{j,T}(k)\}$ given by the kernel weights on the one hand and the recursive weights on the other hand, for $j = 1, 2, \ldots, 5$ with a bandwidth $\mathrm{b}_T = 0.25$. For the kernel weight we took the triangle kernel $K(t) = (1-2|t|)_+$, which satisfies the assumptions of Lemma 4(K-2). The obtained local scalograms $\widehat{\sigma}^2_{j,T}(u)$ of the local wavelet spectrum $\sigma^2_j(u)$, $j = 1, 2, \ldots, 5$, $u \in [0,1]$ are represented in the lower parts of Figures 2 and 3, respectively, with a y-axis in a logarithmic scale. The five corresponding curves exhibit

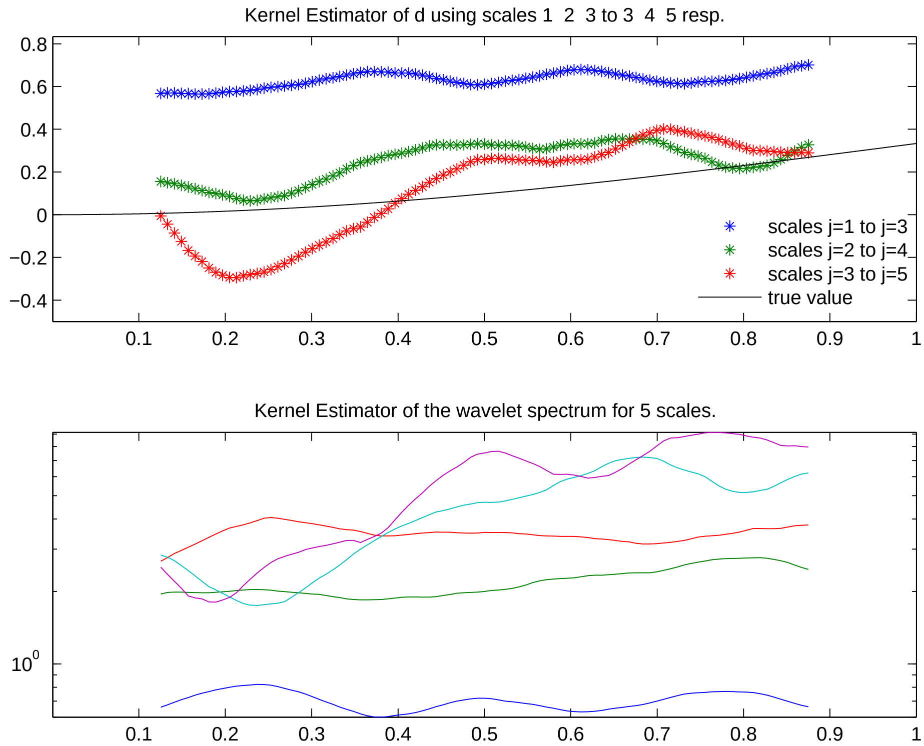

FIGURE 2. Local estimates as functions of $u \in [0,1]$ for the simulated tvARFIMA(1,d,0) using a two-sided triangular kernel. Top: $\hat{d}_T(L;u)$ using scales $j = 1,2,3$ to $3,4,5$ (respectively in blue, green and red) and the true value $d(u)$ (in thin black). Bottom: $\widehat{\sigma}^2_{j,T}(u)$ for $j = 1,\ldots,5$.

different variabilities, the larger j, the larger the variability, which is in accordance with our theoretical findings. On the top of these two figures, we represented the true parameter $d(u)$, $u \in [0,1]$ (plain black) and the corresponding estimators $\widehat{d}_T(u)$ for three sets of scales, namely $j = 1,2,3$ (blue stars), $j = 2,3,4$ (green stars) and $j = 3,4,5$ (red stars), which correspond to $L = 1,2,3$, respectively, and $\ell = 2$ in the three sets of scales. One can observe that the estimates are upper biased for $L = 1,2$, while the estimates is varying more widely for $L = 3$. Again this matches our theoretical findings. One can also observed the difference between the two-side kernel estimator and the recursive estimator. The former exhibits a uniform behavior along time with border effects close to each boundaries of the interval $[0,1]$ (here we dropped the values of $\widehat{d}_T(u)$ for $u < \mathrm{b}_T/2$ and $u > 1-\mathrm{b}_T/2$ to avoid these border effects). In contrast the latter exhibits a diminishing then stabilizing variability along time. Thus it is better adapted for estimating the right part of the interval. It is interesting to note that the choice of L is crucial for this simulated example. This is due to the presence of an autoregressive component leading to a strong positive short-memory autocorrelation with a root close to the unit circle. As a result $d(u)$ is over estimated if a too high frequency band of scales is used (as in the case $L = 1$). On the other hand this bias diminishes drastically as soon as $L \geq 2$. We made similar experiments for a tvARFIMA(0,d,0) process. In this case, this bias is no longer observed for $L = 1$. We have also tried different values

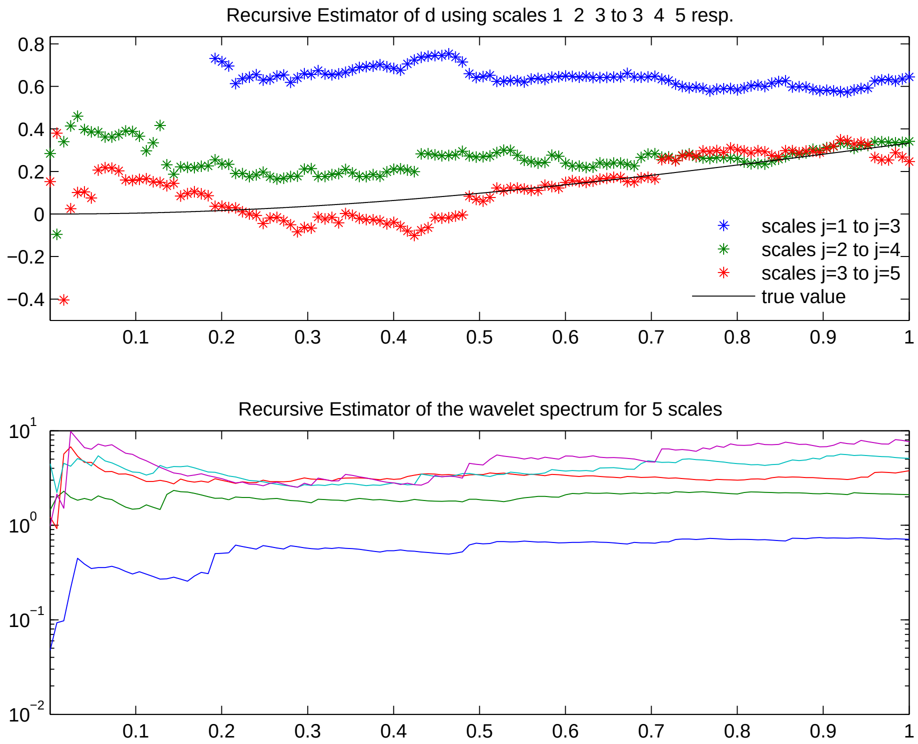

FIGURE 3. Same as Figure 2 using a recursive estimator.

of the bandwidth b_T which also influences the bias and the variability of the estimates in the expected way. Finally we tested our procedures on longer series to check the numerical tractability. The computation of $\widehat{\sigma}^2_{j,T}(u)$ from $X_{1,T}, \ldots, X_{T,T}$, with $T = 2^{15}$ took less than 1 second for the kernel estimator and 7 seconds for the recursive estimator with a 3.00GHz CPU. We note that the recursive version is about ten times slower than the kernel estimator. On the other hand the recursive estimator is adapted to *online* computation, that is, $\widehat{\sigma}^2_{j,T}(t)$ can be computed in a recursive fashion for each new available observation $X_{t,T}$.

5.2. **Real data sets.** We now use real data sets made of a sample of realized log volatility of the YEN versus USD exchange rate between June 1986 and September 2004. The realized log volatility is represented in Figure 4. The series length is $T = 4470$, that is of the same order as the previously simulated series ($T = 2^{12} = 4096$). Viewing the simulated data as a benchmark, we used approximately the same bandwidth parameter $b_T = 0.23$ and the same sets of scales, namely $L = 1, 2, 3$ with $\ell = 3$ in the three cases. The two-sided kernel estimators of the memory parameter are represented in the upper part of Figure 5. As previously we also display the corresponding local scalograms in the lower part of the same figure. We omit the results for the recursive estimator as the non-uniform behavior along time makes the results harder to interpret. One can observe that here as L increases the estimates of $d(u)$ globally increases which may indicate a lower bias at high frequencies. The red curve also varies more widely than the two others, which indicates a poorer reliability. The green curve appears as a good compromise as in the simulated example. It exhibits a 5 years periodic-like behavior, which seems to indicate that the long memory parameter is

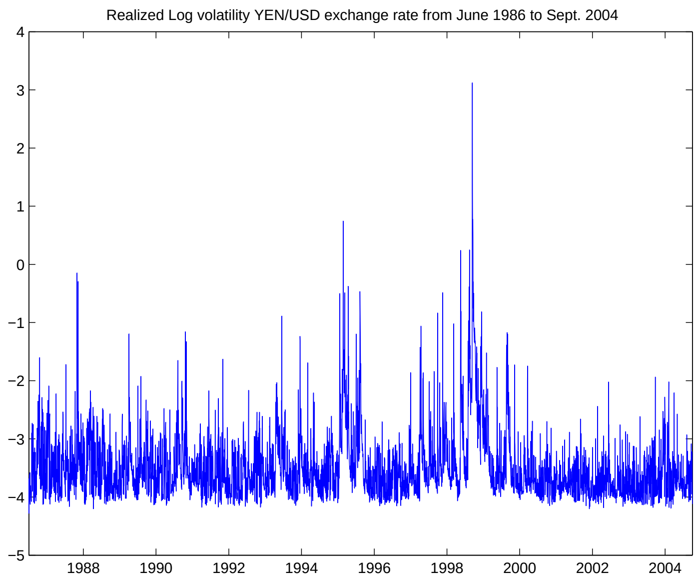

Figure 4. Realized log volatility of the YEN vs USD exchange rate from June 1986 to September 2004.

not constant over time. This seems to be in accordance with the findings of [17] who model long-memory realized volatilities by a change of the model parameters from one regime to another where the different regimes can be explained by the influence of changing market factors (such as the Asian financial crisis of 1998).

6. Conclusion

In this paper we have delivered a semi-parametric, hence fairly general, approach for estimating the time-varying long-memory parameter $d(u)$ of a locally stationary process (or increment process if differencing is necessary beforehand). Apart from modelling the singularity at zero frequency by the curve $d(u)$, we do not need to model the time varying spectrum of the remaining part explicitly. Using a wavelet log-regression estimator, already shown to be well-performing in the stationary situation, continues to work well due to a localization of the wavelet scalograms across time within each scale.

The development of our approach is based on a stationary approximation at each given time point u. As in the stationary case, due to the generality of our semi-parametric spectral density not to be depending on only a finite number of parameters (as in [2], e.g.), we need to concentrate our attention to well estimating around frequency zero (where the amount of the long-memory effect measured by d is visible). So a slightly subtle choice of considered scales for the log-regression has to be done: asymptotically we need that our estimator involves more and more frequencies (i.e. scales) but with a *maximal frequency* tending to zero. In the wavelet domain, this means that the lowest scale used in the estimator will be chosen so that i) the number of wavelet coefficients used in the estimator tends to infinity and ii) this lowest scale itself tends (slowly) to infinity.

Simulations have shown that our estimator performs reasonably well beyond being attractive from the point of view of asymptotic theory. In our real data analysis example, we adopt the approach of [17] and of [6] to assume that realized volatilities of some exchange

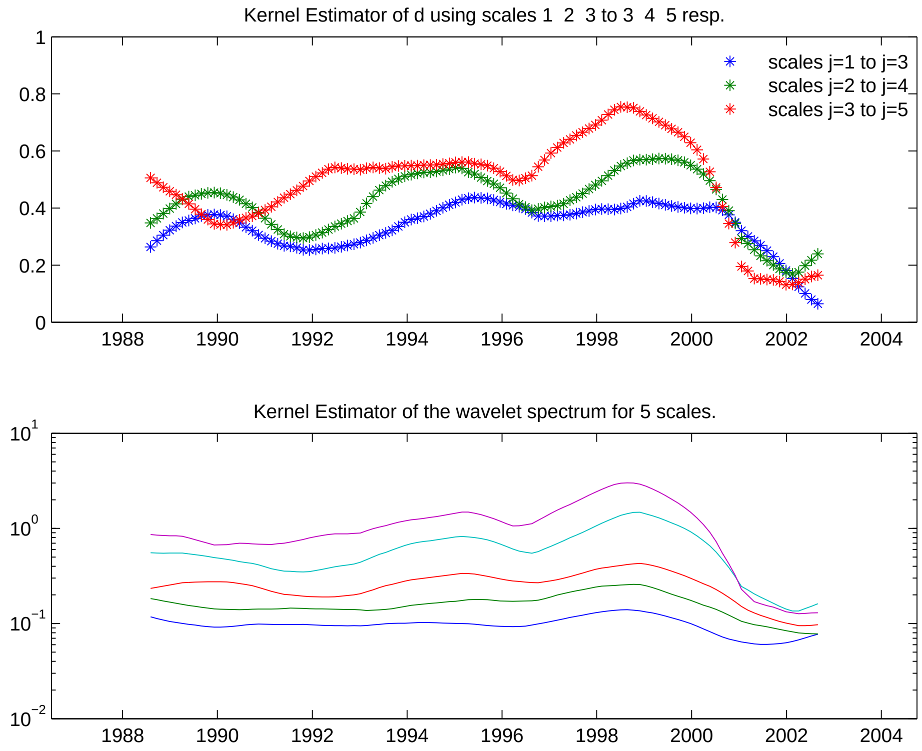

FIGURE 5. Same as Figure 2 for the YEN vs USD exchange rate realized log volatility.

rates follow a long-memory model. We make the interesting observation that for the observed series the long-memory parameter can clearly not be considered to be constant over time - which suggests that in explaining the persistent correlation in this exchange data there are certainly periods of stronger persistence followed by periods of weaker persistence.

ACKNOWLEDGEMENTS

We acknowledge financial support from the TSIFIQ project of the “Fondation Telecom”, from the IAP research network grant no P6/03 of the Belgian government (Belgian Science Policy), and from the “Projet d’Actions de Recherche Concertées” no 07/12-002 of the “Communauté française de Belgique”, granted by the “Académie universitaire Louvain”.

We also thank S.-H. Wang and L. Bauwens for providing the data example of Section 5.2 and for their helpful comments related to this analysis.

APPENDIX A. POSTPONED PROOFS

of Proposition 1. By [12, Proposition 3], there is a constant C_1 such that, for all $j \geq 0$ and all $\lambda \in [-\pi, \pi]$,

$$|H_j(\lambda)| \leq C_3\, 2^{j/2}\, |2^j \lambda|^M \,(1 + 2^j |\lambda|)^{-\alpha-M} \ . \tag{49}$$

Applying (1), (7), (13) and (17), we get, for any $u \in \mathbb{R}$, $j \geq 0$ and $k \in \{0, \ldots, T_j - 1\}$,

$$W_{j,k;T} = W_{j,k}(u) + R_{j,k}(u;T) \ , \tag{50}$$

where

$$R_{j,k}(u;T) = \int_{-\pi}^{\pi} \sum_{s\in\mathbb{Z}} \tilde{h}_{j,s} \left[A^0_{2^j k-s,T}(\lambda) - A(u;\lambda)\right] \mathrm{e}^{\mathrm{i}\lambda(2^j k-s)} \, \mathrm{d}Z(\lambda) \; .$$

The main approximation result consists in bounding

$$S_j(u;T) = \sum_{k=0}^{T_j-1} \gamma_{j,T}(k) R^2_{j,k}(u;T)$$

and

$$D_j(u;T) = \sum_{k=0}^{T_j-1} \gamma_{j,T}(k) W_{j,k}(u) R_{j,k}(u;T) \; .$$

In the following C denotes some multiplicative constant. Using (2), (3), and (60) in Lemma 1, we have

$$\left|\sum_{s\in\mathbb{Z}} \tilde{h}_{j,s} \left[A^0_{2^j k-s,T}(\lambda) - A(u;\lambda)\right] \mathrm{e}^{\mathrm{i}\lambda(2^j k-s)}\right| \leq C\; 2^{jp}\; |\lambda|^{-D} \; \left\{2^{j/2}\left|2^j k/T - u\right| + 2^{3j/2}/T\right\} \; .$$

Recall that D denotes an exponent less than $1/2$ which appears in the Conditions (2) and (3). Since $D < 1/2$, we get

$$\mathbb{E}\left[R^2_{j,k}(u;T)\right] \leq C\; 2^{2jp}\; 2^{3j}\; T^{-2} \; \left\{1 + (k - Tu2^{-j})^2\right\} \; .$$

Since we assumed $\alpha > 1/2 - d(u)$, we can take D large enough so that $1-\alpha-d(u) < D < 1/2$ (by adapting the constant c appearing in the afore mentioned conditions). Hence we can assume in the following that

$$M > d(u) - 1/2 \quad \text{and} \quad d(u) + D + \alpha > 1 \; . \tag{51}$$

By (18) we also obtain that

$$\left|\mathbb{E}\left[W_{j,k}(u) R_{j,k}(u;T)\right]\right| \leq C\; 2^{jp} \; \left\{2^{j/2}\left|2^j k/T - u\right| + 2^{3j/2}/T\right\} \int_{-\pi}^{\pi} \left|\tilde{H}_j(\lambda) A(u;\lambda)\right| \; |\lambda|^{-D} \, \mathrm{d}\lambda$$

Using (16), (4), (5), $f^*(u,\lambda) \leq C f^*(u,0)$ (by (6)), and (49), we further have

$$\begin{aligned} \int_{-\pi}^{\pi} \left|\tilde{H}_j(\lambda) A(u;\lambda)\right| \; |\lambda|^{-D} \, \mathrm{d}\lambda &\leq C \int_{-\pi}^{\pi} |H_j(\lambda)| \; \sqrt{f(u,\lambda)}\; |\lambda|^{-D} \, \mathrm{d}\lambda \\ &\leq C\; \sqrt{f^*(u,0)}\; 2^{j(d(u)+D-1/2)} \; , \end{aligned}$$

where we used that $\int_{\mathbb{R}} |\xi|^{M-d(u)-D} \; (1+|\xi|)^{-\alpha-M} \mathrm{d}(\xi) < \infty$ by (51). The last displays provide simple bounds for the expectations of S_j and D_j. To bound their variance, we rely on the Gaussian assumption, which implies

$$\mathrm{Cov}\left(R^2_{j,k}(u;T), R^2_{j,k'}(u;T)\right) = 2\mathrm{Cov}^2\left(R_{j,k}(u;T), R_{j,k'}(u;T)\right)$$

and

$$\begin{aligned} &\mathrm{Cov}\left(W_{j,k}(u) R_{j,k}(u;T), W_{j,k'}(u) R_{j,k'}(u;T)\right) \\ &\qquad = \mathrm{Cov}\left(W_{j,k}(u), W_{j,k'}(u)\right) \mathrm{Cov}\left(R_{j,k}(u;T), R_{j,k'}(u;T)\right) \\ &\qquad\qquad + \mathrm{Cov}\left(R_{j,k}(u;T), W_{j,k'}(u)\right) \mathrm{Cov}\left(W_{j,k}(u), R_{j,k'}(u;T)\right) \; . \end{aligned}$$

Let us first provide bounds of $\mathbb{E}\left[R_{j,k}(u;T)R_{j,k'}(u;T)\right]$ and $\mathbb{E}\left[W_{j,k}(u)R_{j,k'}(u;T)\right]$ for $k,k'=0,\dots,T_j-1$. Proceeding as previously, under the condition (51), we get (in fact the cases above $k=k'$ are particular cases)

$$\left|\mathbb{E}\left[R_{j,k}(u;T)R_{j,k'}(u;T)\right]\right| \le C\ 2^{2jp}\ 2^{3j}\ T^{-2}\ \left\{1+|k-Tu2^{-j}|\right\}\left\{1+|k'-Tu2^{-j}|\right\}\ .$$

and

$$\left|\mathbb{E}\left[W_{j,k}(u)R_{j,k'}(u;T)\right]\right| \le C\ 2^{jp}\ \sqrt{f^*(u;0)}\ 2^{j(d(u)+D-1/2)}\ \left\{2^{j/2}\left|2^jk'/T-u\right|+2^{3j/2}/T\right\}\ .$$

We obtain the same bound for $\mathrm{Var}^{1/2}\left(S_j(u;T)\right)$ and $\mathbb{E}\left[S_j(u;T)\right]$ and thus, using the definition of Γ in (22),

$$\left|\mathbb{E}\left[S_j^2(u;T)\right]\right|^{1/2} \le C\ 2^{2jp}\ 2^{3j}\ T^{-2}\ \left\{\Gamma_0(u;j,T)+\Gamma_2(u;j,T)\right\}\ . \tag{52}$$

For $D_j(u;T)$, we obtain

$$\left|\mathbb{E}\left[D_j(u;T)\right]\right| \le C\ 2^{jp}\ \sqrt{f^*(u;0)}\ 2^{j(d(u)+D-1/2)}\ 2^{3j/2}\ T^{-1}\ \left\{\Gamma_0(u;j,T)+\Gamma_1(u;j,T)\right\}\ .$$

Denote by $B_j(u)$ the variance of the (stationary) process $\{W_{j,k}(u),\ k\in\mathbb{Z}\}$. We then obtain that $\mathrm{Var}^{1/2}\left(D_j(u;T)\right)$ is at most

$$C\ 2^{jp}\ 2^{3j/2}\ T^{-1}\ \left\{\Gamma_0(u;j,T)+\Gamma_1(u;j,T)\right\}\left\{B_j^{1/2}(u)+\sqrt{f^*(u;0)}\ 2^{j(d(u)+D-1/2)}\right\}\ .$$

Observe that by [12, Theorem 1] we have, since $M > d(u)-1/2$ and $\alpha > 1/2-d(u)$, $B_j(u) \le C\ f^*(u;0)\ 2^{2d(u)j}$. Hence, since $D<1/2$,

$$\left|\mathbb{E}\left[D_j^2(u;T)\right]\right|^{1/2} \le C\ 2^{jp}\ \sqrt{f^*(u;0)}\ 2^{j(3/2+d(u))}\ T^{-1}\ \left\{\Gamma_0(u;j,T)+\Gamma_1(u;j,T)\right\}\ . \tag{53}$$

By (20) and (50), we have

$$\widehat{\sigma}^2_{j,T}(u) = \widetilde{\sigma}^2_{j,T}(u) + S_j(u;T) + D_j(u;T)\ , \tag{54}$$

where $\widetilde{\sigma}^2_{j,T}(u)$ is defined in (23). The bound (35) now follows from (52), (53), (54) and Assumption 1 (iv). □

of Theorem 2. The proof follows the lines of [13, Theorem 2], in which the stationary case is considered, *i.e.* $\gamma_{j,T}(u)=1$. We first observe that, for any $\mu=[\mu_0\ \dots\ \mu_\ell]^T\in\mathbb{R}^{\ell+1}$, we may write

$$\mu^T\widetilde{S}_L(u) = \xi_L^T\Delta_L\xi_L\ ,$$

where ξ_L is a Gaussian vector with entries $\left(W_{L+i,k}(u)\right)_{0\le i\le\ell,\,0\le k\le T_{L+i}}$ and Δ_L is the diagonal matrix with diagonal entries $\left(2^{-2Ld(u)}\delta_{L,T}^{-1/2}\mu_i\gamma_{L+i,T}(u)\right)_{0\le i\le\ell,\,0\le k\le T_{L+i}}$. We may thus apply [14, Lemma 12].

To obtain (40), it is thus sufficient to show that

$$\rho(\Delta_L)\rho\left(\mathrm{Cov}(\xi_L)\right)\to 0\ , \tag{55}$$

where $\rho(A)$ denotes the spectral radius of A, and

$$\mathrm{Cov}(\mu^T\widetilde{S}_L(u)) \to (f^*(u,0))^2\mu^T\Sigma\mu\ . \tag{56}$$

We have, by (28) and Assumption 1(i),

$$\rho(\Delta_L) \le 2^{-2Ld(u)}\delta_{L,T}^{-1/2}\max_{0\le i\le\ell}|\mu_i|\max_{0\le i\le\ell}\delta_{L+i,T} = o\left(2^{-2Ld(u)}\right)\ .$$

Using [13, Lemma 6], [14, Lemma 11] and that $\mathbf{D}_{L+i}$ is the spectral density of the process $\{W_{L+i,k}(u),\ k \in \mathbb{Z}\}$, we have

$$\rho\left(\mathrm{Cov}(\xi_L)\right) \leq \sum_{i=0}^{\ell} \rho\left(\mathrm{Cov}([W_{L+i,k}(u),\ k=0,\dots,T_{L+i}])\right) \leq 2\pi \sum_{i=0}^{\ell} \|\mathbf{D}_{L+i}\|_\infty .$$

By [12, Theorem 1], since we assumed $M \geq d(u)$ and $\alpha > 1/2 - d(u)$, we have $\|\mathbf{D}_{L+i}\|_\infty = O(2^{2Ld(u)})$. This with the last two displays implies (55).

We now compute the asymptotic covariance matrix of $\widetilde{S}_L(u)$. Let $0 \leq j' \leq j$. Using (27) and the Gaussian assumption, we have

$$\mathrm{Cov}\left(\widetilde{\sigma}^2_{j,T}(u), \widetilde{\sigma}^2_{j',T}(u)\right) = \sum_{k=0}^{T_j-1}\sum_{k'=0}^{T_{j'}-1} \gamma_{j,T}(k)\gamma_{j',T}(k')\mathrm{Cov}\left(W^2_{j,k}(u), W^2_{j',k'}(u)\right)$$

$$= 2\sum_{v=0}^{2^{j-j'}-1}\sum_{k=0}^{T_j-1}\sum_{l\in\mathcal{T}_j(j-j',v)} \gamma_{j,T}(k)\gamma_{j',T}(l2^{j-j'}+v)\mathrm{Cov}^2\left(W_{j,k}(u), W_{j',l2^{j-j'}+v}(u)\right) .$$

Using [12, Corollary 1], we have

$$\mathrm{Cov}\left(W_{j,k}(u), W_{j',l2^{j-j'}+v}(u)\right) = \int_{-\pi}^{\pi} \mathbf{D}_{j,j-j',v}(\lambda)\mathrm{e}^{\mathrm{i}\lambda(k-l)}\,\mathrm{d}\lambda ,$$

where $\mathbf{D}_{j,j-j'} = [\mathbf{D}_{j,j-j',v}]_{v=0,\dots,2^{j-j'}-1}$ denotes the $2^{j-j'}$-dimensional cross-spectral density between $W_{j,k}(u)$ and $[W_{j',l2^{j-j'}+v}(u)]_{v=0,\dots,2^{j-j'}-1}$. It follows from the last two displays and (26) that

$$\mathrm{Cov}\left(\widetilde{\sigma}^2_{j,T}(u), \widetilde{\sigma}^2_{j',T}(u)\right) = 2\sum_{v=0}^{2^{j-j'}-1}\int_{-\pi}^{\pi} \Phi_{j,T}(\lambda;0,0)\overline{\Phi}_{j,T}(\lambda;j-j',v)\ \widetilde{\mathbf{D}}_{j,j-j',v}(\lambda)\,\mathrm{d}\lambda ,$$

where

$$\widetilde{\mathbf{D}}_{j,j-j',v}(\lambda) = \int_{-\pi}^{\pi} \mathbf{D}_{j,j-j',v}(\xi)\overline{\mathbf{D}}_{j,j-j',v}(\xi-\lambda)\mathrm{d}\xi .$$

By [12, Theorem 1(b)], since we assumed $M \geq d(u)$ and $\alpha > 1/2 - d(u)$, using (6), we have, for $j = L+i$ and $j' = L+i'$ with $i' \leq i$ fixed,

$$\|2^{-2d(u)j}\mathbf{D}_{j,j-j'} - f^*(u,0)\mathbf{D}_{\infty,i-i'}(\cdot;d(u))\|_\infty \to 0 .$$

The last three displays, (42), Lemma 2 and Assumption 1 yield

$$\mathrm{Cov}\left(\widetilde{S}_{L,T}(u)\right) \to (f^*(u,0))^2\Sigma ,$$

and hence (56). □

of Theorem 1. By (19) and (21),

$$\mathbb{E}\left[\widetilde{\sigma}^2_{j,T}(u)\right] = \mathbb{E}\left[W^2_{j,k}(u)\right] = \sigma^2_j(u) . \tag{57}$$

Since the wavelet coefficients (17) are those of a stationary process, their behavior at large scales ($j \to \infty$) can be studied using [12, Theorem 1]. By [12, Theorem 1], since we assumed (6) and $M > d(u) - 1/2$ and $\alpha > (1+\beta)/2 - d(u)$, we obtain (37). In the following we denote

$$K^*_u = f^*(u,0)\kappa(d(u)) .$$

We now provide a bound for

$$\mathrm{Var}\left(\widetilde{\sigma}^2_{j,T}(u)\right) = \sum_{k,k'=0}^{T_j-1} \gamma_{j,T}(k)\gamma_{j,T}(k')\mathrm{Cov}\left(W^2_{j,k}(u), W^2_{j,k'}(u)\right)$$
$$= 2\int_{-\pi}^{\pi} |\Phi_{j,T}(\lambda;0,0)|^2 \ \mathbf{D}_j^{\star 2}(\lambda)\ \mathrm{d}\lambda\ ,$$

where $\mathbf{D}_j$ denotes the spectral density of the stationary Gaussian process $\{W_{j,k}\ , k\in\mathbb{Z}\}$, $\Phi_{j,T}$ is defined in (26) and, for any (2π)-periodic function g, $g^{\star 2} = g\star g(\lambda) = \int_{-\pi}^{\pi} g(\lambda - \xi)g(\xi)\mathrm{d}\xi$.

By [12, Theorem 1] we have, since $M \geq d(u)$ and $\alpha > 1/2 - d(u)$,

$$\int_{-\pi}^{\pi} |\mathbf{D}_j(\lambda)|^2 \ \mathrm{d}\lambda = O\left(2^{4jd(u)}\right)\ .$$

(the constants depend on $f^*(u;0)$ only). Using the last two displays and Assumption 1(ii) with $i = i' = v = v' = 0$, we get that

$$\mathrm{Var}\left(\widetilde{\sigma}^2_{j,T}(u)\right) = O(2^{4jd(u)}\delta_{j,T})\ . \tag{58}$$

Using (54) and (57), $\mathbb{E}\left[(\widehat{\sigma}^2_{j,T}(u) - K^*_u\, 2^{2jd(u)})^2\right]$ is at most

$$C\ \left\{\mathrm{Var}\left(\widetilde{\sigma}^2_{j,T}(u)\right) + \mathbb{E}\left[S_j^2(u;T)\right] + \mathbb{E}\left[D_j^2(u;T)\right]\right\} + O\left(2^{2(2d(u)-\beta)j}\right)$$
$$= O\left(2^{4jd(u)}\delta_{j,T} + 2^{(6+4p)j}T^{-4}\delta_{j,T}^{-4} + 2^{(3+2p+2d(u))j}T^{-2}\delta_{j,T}^{-2} + 2^{2(2d(u)-\beta)j}\right)\ .$$

where we used (58), (52), (53) and (30). Using (38), the last display gives (39). □

of Theorem 3. We first show that

$$\delta_{L,T}^{-1/2}\left(2^{-2Ld(u)}\begin{bmatrix}\widehat{\sigma}^2_{L,T}(u)\\ \widehat{\sigma}^2_{L+1,T}(u)\\ \vdots\\ \widehat{\sigma}^2_{L+\ell,T}(u)\end{bmatrix} - K^*_u\begin{bmatrix}1\\ 2^{2d(u)}\\ \vdots\\ 2^{2\ell d(u)}\end{bmatrix}\right) \Rightarrow \mathcal{N}\left(0, (f^*(u,0))^2\Sigma(u)\right)\ . \tag{59}$$

Observe that the weak convergence (59) is the same as (44) except for the centering term. Relation (44) is valid since the assumptions of Corollary 1 hold. Applying $\delta_{L,T}\to 0$, Proposition 1 and the left-hand side condition of (47), we have that, for any $j = L+i$ with a fixed $i = 0,\ldots,\ell$,

$$\delta_{L,T}^{-1/2}2^{-2Ld(u)}\mathbb{E}\left[\widehat{\sigma}^2_{j,T}(u)\right] = \delta_{L,T}^{-1/2}2^{-2Ld(u)}\mathbb{E}\left[\widetilde{\sigma}^2_{j,T}(u)\right] + o(1)\ .$$

The bias control (37) and the right-hand side condition of (47) then imply

$$\delta_{L,T}^{-1/2}2^{-2Ld(u)}\mathbb{E}\left[\widehat{\sigma}^2_{j,T}(u)\right] = \delta_{L,T}^{-1/2}f^*(u,0)\kappa(d(u))2^{2id(u)} + o(1)\ .$$

This, with (44) gives the weak convergence (59) .

The convergence (48) now follows from (59) by applying the δ-method as in [13, Proposition 3]. Indeed, define

$$g(x) = \sum_{i=0}^{\ell} w_i \log(x_i) \quad \text{for all} \quad x = [x_0\ \ldots\ x_\ell]^T\ .$$

Observe that, by (24) and (25), we have

$$g\left(2^{-2Ld(u)}[\widehat{\sigma}^2_{L,T}(u)\ \widehat{\sigma}^2_{L+1,T}(u)\ \dots\ \widehat{\sigma}^2_{L+\ell,T}(u)]^T\right) = \widehat{d}_T(L)$$

and

$$g\left(f^*(u,0)\kappa(d(u))[1\ 2^{2d(u)}\ \dots\ 2^{2\ell d(u)}]^T\right) = d(u)\ .$$

Thus (48) follows from (59) by computing the gradient of g at the centering term,

$$\nabla g\left(f^*(u,0)\kappa(d(u))[1\ 2^{2d(u)}\ \dots\ 2^{2\ell d(u)}]^T\right) = \frac{[w_0\ w_1 2^{-2d(u)}\ \dots\ w_\ell 2^{-2\ell d(u)}]^T}{f^*(u,0)\kappa(d(u))}\ .$$

□

Appendix B. Technical lemmas

Lemma 1. *Assume (W-1)–(W-4). Let $h_{j,\cdot}$ the wavelet detail filter at scale index j and $\tilde{h}_{j,\cdot}$ any factorization of it by $\mathbf{\Delta}^p$ with $p\in\{0,\dots,M\}$. Then we have*

$$\sum_{s\in\mathbb{Z}} |\tilde{h}_{j,s}| \leq C\ 2^{j(p+1/2)} \quad \text{and} \quad \sum_{s\in\mathbb{Z}} (1+|s|)\ |\tilde{h}_{j,s}| \leq C\ 2^{j(p+3/2)}\ . \tag{60}$$

Lemma 2. *Suppose Assumption 1 holds. Let $i,i'\geq 0$, $v\in\{0,\dots,2^i-1\}$ and $v'\in\{0,\dots,2^{i'}-1\}$. Define, for any (2π)-periodic function g,*

$$I_T(g) = \delta_{j,T}^{-1}\int_{-\pi}^{\pi} \Phi_{j,T}(\lambda;i,v)\overline{\Phi_{j,T}(\lambda;i',v')}\ g(\lambda)\,\mathrm{d}\lambda\ .$$

Then the two following assertions hold.

(i) If $h\to g$ in $L^\infty([-\pi,\pi])$, then $\sup_{T\geq 0}|I_T(h)-I_T(g)|\to 0$.

(ii) If $g\in L^\infty([-\pi,\pi])$ is continuous at zero, then, as $T\to\infty$, $I_T(g)\to V(i,v;i',v')\,g(0)$.

Proof. By linearity of I_T, we may take $g=0$ to prove Assertion (i). We have, by the Cauchy-Schwarz inequality

$$|I_T(h)| \leq \|h\|_\infty \left[\delta_{j,T}^{-1/2}\|\Phi_{j,T}(\cdot;i,v)\|_2\right]\ \left[\delta_{j,T}^{-1/2}\|\Phi_{j,T}(\cdot;i',v')\|_2\right]\ .$$

Using Assumption 1(ii), the terms between brackets are bounded independently of j and we obtain (i).

We now prove (ii). By linearity of I_T, we may assume $g(0)=1$. By Assumption 1(ii), we have $I_T(1)\to V(i,v;i',v')$. On the other hand, we have, for any $\eta>0$

$$\begin{aligned}|I_T(g)-I_T(1)| &= \left|I_T((g-1)\mathbb{1}_{[-\eta,\eta]} + (g-1)\mathbb{1}_{[-\eta,\eta]^c})\right| \\ &\leq \left|I_T((g-1)\mathbb{1}_{[-\eta,\eta]})\right| + \left|I_T((g-1)\mathbb{1}_{[-\eta,\eta]^c})\right|\ .\end{aligned}$$

Observe that by continuity of g at the origin, $\|(g-1)\mathbb{1}_{[-\eta,\eta]}\|_\infty\to 0$ as $\eta\to\infty$. By (i), we get $\left|I_T((g-1)\mathbb{1}_{[-\eta,\eta]})\right|\to 0$ as $\eta\to\infty$. It thus only remains to show that $\left|I_T((g-1)\mathbb{1}_{[-\eta,\eta]^c})\right|\to 0$ for any $\eta>0$. This follows from the bound

$$\left|I_T((g-1)\mathbb{1}_{[-\eta,\eta]^c})\right| \leq \|g-1\|_1 \left[\delta_{j,T}^{-1/2}\sup_{\eta\leq|\lambda|\leq\pi}|\Phi_{j,T}(\lambda;i,v)|\right]\left[\delta_{j,T}^{-1/2}\sup_{\eta\leq|\lambda|\leq\pi}\left|\Phi_{j,T}(\lambda;i',v')\right|\right]\ ,$$

and by applying Assumption 1(iii). □

Lemma 3. *For any $a > 0$ and $b > 0$, there exists $c > 0$ such that*

$$|z^\alpha - 1| \le c\,\{1 + \log(|z|)\}\,\alpha \quad \text{for all} \quad \alpha \in [0, a],\ z \in \mathbb{C} \quad \text{with} \quad |z| \le b\ .$$

Lemma 4. *Assume one of the following.*

(K-1) $K = \mathbb{1}_{[-1/2,1/2]}$.

(K-2) K *is compactly supported and* $|\hat{K}(\xi)| = o(|\xi|^{-3/2})$ *as* $|\xi| \to \infty$, *where* $\hat{K}$ *denotes the Fourier transform of* K.

(K-3) K *is integrable,* $\hat{K}$ *has an exponential decay,* i.e. *for some* $c > 0$, $|\hat{K}(\xi)| = O(\exp(-c|\xi|))$ *as* $|\xi| \to \infty$, $K(t) = O(|t|^{-p_0})$ *as* $|t| \to \infty$ *for some* $p_0 > 3$, *the derivative* K' *of* K *satisfies* $|K'(t)| = O(|t|^{-p_1})$ *as* $|t| \to \infty$ *for some* $p_1 > 1$ *and* $T_j \exp(-c'\mathrm{b}_T T_j) = O(1)$ *for any* $c' > 0$.

Suppose that $\mathrm{b}_T \to 0$ *and that* j *depends on* T *so that* $T_j \mathrm{b}_T \to \infty$ *as* $T \to \infty$. *Then, for weights given by (32), Assumption 1 is satisfied with*

$$\delta_{j,T} \sim \frac{\|K\|_\infty}{\|K\|_1} (\mathrm{b}_T T_j)^{-1} \tag{61}$$

$$V(i, v; i', v') = 2\pi \frac{\|K\|_2^2}{\|K\|_1 \|K\|_\infty}\, 2^{-i-i'}\ , \quad i, i' \ge 0,\ 0 \le v < 2^i,\ 0 \le v' < 2^{i'}\ . \tag{62}$$

Proof. For convenience, we will omit the subscripts ${}_T$ and ${}_{j,T}$ in this proof section when no ambiguity arises. Under (K-1), one has $\rho = \mathrm{b}T_j + O(1)$. Under (K-2), K is uniformly continuous on its compact support S and, since $u \in (0, 1)$, $\mathrm{b} \to 0$ and $T_j \mathrm{b} \to \infty$, S eventually falls between the extremal points of $\{(uT_j - k)/(\mathrm{b}T_j),\ k = 0, \dots, T_j - 1\}$. Thus,

$$(\mathrm{b}T_j)^{-1} \sum_{k=0}^{T_j-1} K((uT_j - k)/(\mathrm{b}T_j)) \to \int_S K(s)\, \mathrm{d}s = \|K\|_1\ .$$

Under (K-3), using that $|K'(t)| \le c(1 + |t|)^{-p_1}$ for some $p_1 > 1$ and $c > 0$, we get

$$(\mathrm{b}T_j)^{-1} \sum_{k=0}^{T_j-1} K((uT_j - k)/(\mathrm{b}T_j)) - \int_{(u-1)/\mathrm{b}}^{u/\mathrm{b}} K(s)\, \mathrm{d}s = O\left((\mathrm{b}T_j)^{-2} \sum_{l=0}^{T_j} (1 + l/(\mathrm{b}T_j))^{-p_1}\right)$$
$$= O\left((\mathrm{b}T_j)^{-1}\right)\ .$$

Hence the last three displays yield that, in all cases,

$$\rho_{j,T} \sim \|K\|_1 (\mathrm{b}_T T_j)\ . \tag{63}$$

The asymptotic equivalence (61) then follows from the definitions (32) and (28), and we obtain Assumption 1(i) by (14).

Let us now prove that Assumption 1(ii) holds under (K-1), (K-2) and (K-3), successively. Note that, by definition of (26), we have

$$\int_{-\pi}^{\pi} \Phi_{j,T}(\lambda; i, v) \overline{\Phi_{j,T}(\lambda; i', v')}\, \mathrm{d}\lambda = 2\pi \sum_{l \in \mathcal{T}_j(i,v) \cap \mathcal{T}_j(i',v')} \gamma_{j-i,T}(2^i l + v) \gamma_{j-i',T}(2^{i'} l + v')\ . \tag{64}$$

Under (K-1), using $2^{-i} T_{j-i} \sim 2^{-i'} T_{j-i'} \sim T_j$ by (28), $\mathrm{b}T_j \to \infty$ and $\mathrm{b} \to 0$, we easily get that the supports of the sequences $\{\gamma_{j-i,T}(2^i l + v),\ l \ge 0\}$ and $\{\gamma_{j-i',T}(2^{i'} l + v),\ l \ge 0\}$ are eventually included in $\mathcal{T}_j(i, v) \cap \mathcal{T}_j(i', v')$ and their intersection is of length asymptotically equivalent to $\mathrm{b}T_j$. Hence, using (64), (61) and (63) with $\|K\|_1 = \|K\|_\infty = 1$, we obtain that, in this case,

$$\delta^{-1} \int_{-\pi}^{\pi} \Phi_{j,T}(\lambda; i, v) \overline{\Phi_{j,T}(\lambda; i', v')} \, \mathrm{d}\lambda \sim 2\pi \, \frac{T_j^2}{T_{j-i} T_{j-i'}} \; .$$

By (14), this is Assumption 1(ii) with $V(i, v; i', v') = 2\pi 2^{-i-i'}$ which coincides with (62) under (K-1).

Under (K-2), we proceed by interpreting the sum in (64) as a Riemann approximation of $\int K^2$ up to a normalization factor. For $l \in \mathcal{T}_j(i, v) \cap \mathcal{T}_j(i', v')$, we approximate

$$\begin{aligned} J_l &= (\mathrm{b}T_j)^{-1} \rho_{j-i,T} \rho_{j-i',T} \gamma_{j-i,T}(2^i l + v) \gamma_{j-i',T}(2^{i'} l + v') \\ &= (\mathrm{b}T_j)^{-1} \, K(\{uT_{j-i} - (2^i l + v)\}/\{\mathrm{b}T_{j-i}\}) \, K(\{uT_{j-i'} - (2^{i'} l + v')\}/\{\mathrm{b}T_{j-i'}\}) \; , \end{aligned}$$

by the local average

$$\tilde{J}_l = \int_{I_l} K^2(s) \, \mathrm{d}s \; ,$$

where I_l is defined as the interval $[\{uT_j - (l+1)\}/\{\mathrm{b}T_j\}, \{uT_j - l\}/\{\mathrm{b}T_j\}]$. Observe that

$$\sup_{s \in I_l} \left| s - \{uT_{j-i} - (2^i l + v)\}/\{\mathrm{b}T_{j-i}\} \right| \le \frac{1}{\mathrm{b}T_j} + \left| \frac{1}{\mathrm{b}T_j} - \frac{2^i}{\mathrm{b}T_{j-i}} \right| |l - uT_j| + u \frac{|T_{j-i} - 2^i T_j|}{\mathrm{b}T_{j-i}} \frac{|v|}{\mathrm{b}T_{j-i}} \; .$$

Using (14), $i, v = O(1)$ and $l = O(T_j)$, we obtain, for any fixed integers i and v,

$$\sup_{0 \le l \le 2T_j} \sup_{s \in I_l} \left| s - \{uT_{j-i} - (2^i l + v)\}/\{\mathrm{b}T_{j-i}\} \right| = O((\mathrm{b}T_j)^{-1}) \; , \tag{65}$$

and the same holds if i, v is replaced by i', v'. Note that

$$\mathcal{T}_j(i, v) \cap \mathcal{T}_j(i', v') = \{0, 1 \ldots, \{2^{-i}(T_{j-i} - v)\} \wedge \{2^{-i'}(T_{j-i'} - v)\} - 1\} \; ,$$

which, by (14) and the fact that K is compactly supported, is eventually contained in $\{0, 1, \ldots, 2T_j\}$ and eventually contains the set of l's such that $\tilde{J}_l \neq 0$, which is of size $O(\mathrm{b}T_j)$. By (65), we also see that, out of a set of length $O(\mathrm{b}T_j)$, both J_l and $\tilde{J}_l$ vanish. Hence we have

$$\sum_{l \in \mathcal{T}_j(i,v) \cap \mathcal{T}_j(i',v')} \left| J_l - \tilde{J}_l \right| = O\left(\mathrm{b}T_j \sup_l |J_l - \tilde{J}_l| \right) \; .$$

Using (65) and the uniform continuity of K, there exists a constant c such that

$$\sup_l |J_l - \tilde{J}_l| \le (\mathrm{b}T_j)^{-1} \sup_{|s-t|, |s-t'| \le c/(\mathrm{b}T_j)} \left| K^2(s) - K(t)K(t') \right| = o((\mathrm{b}T_j)^{-1}) \; .$$

The last two displays, (64) and the definitions of J_l and $\tilde{J}_l$ thus yield

$$\frac{\rho_{j-i,T} \rho_{j-i',T}}{\mathrm{b}T_j} \int_{-\pi}^{\pi} \Phi_{j,T}(\lambda; i, v) \overline{\Phi_{j,T}(\lambda; i', v')} \, \mathrm{d}\lambda \sim 2\pi \int K^2(s) \, \mathrm{d}s = 2\pi \|K\|_2^2 \; . \tag{66}$$

By (61) and (63), this gives Assumption 1(ii) with $V(i, v; i', v')$ given by (62).

Under (K-3), we proceed similarly but we can no longer use that K has a compact support. Instead we use that K is bounded and $|K'(t)| \le c'(3 + |t|)^{-p_1}$ for some $p_1 > 1$ and $c' > 0$ and thus, for any $c > 0$, as soon as $(c+1)/(\mathrm{b}T_j) \le 1$,

$$\sup_{s \in I_l} \sup_{|t-s|, |t'-s| \le c/(\mathrm{b}T_j)} \left| K^2(s) - K(t)K(t') \right| \le c'' \, (\mathrm{b}T_j)^{-1} (2 + |uT_j - l|/(\mathrm{b}T_j))^{-p} \; .$$

With (65) and since the length of $\mathcal{T}_j(i,v) \cap \mathcal{T}_j(i',v')$ is $O(T_j)$, we get

$$\sum_{l \in \mathcal{T}_j(i,v) \cap \mathcal{T}_j(i',v')} \left| J_l - \tilde{J}_l \right| = O\left((\mathrm{b}T_j)^{-2} \sum_{k=0}^{T_j} (1 + k/(\mathrm{b}T_j))^{-p} \right) = O\left((\mathrm{b}T_j)^{-1} \right) .$$

Moreover

$$\sum_{l \in \mathcal{T}_j(i,v) \cap \mathcal{T}_j(i',v')} \tilde{J}_l = \int_{-u'/\mathrm{b}}^{u/\mathrm{b}} K^2(s) \, \mathrm{d}s \to \|K\|_2^2 ,$$

where $u' = [\{2^{-i}(T_{j-i} - v)\} \wedge \{2^{-i'}(T_{j-i'} - v')\}]/T_j - u \to 1 - u$ by (14). This yields (66) as in the previous case and thus the same conclusion holds.

Let us now show that Assumption 1 (iii) holds under (K-1), (K-2) and (K-3), successively. Under (K-1), we have

$$|\Phi(\lambda; i, v)| = \rho_{j-i,T}^{-1} \left| \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i}k\lambda} \right| ,$$

where $N = N_{j,T}$ denotes the number of $l \in \mathcal{T}_j(i,v)$ such that $\gamma_{j-i,T}(2^i l + v) > 0$. Since the Dirichlet kernel satisfies

$$|D_N(\lambda)| = \left| \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i}k\lambda} \right| = \left| \frac{\sin(\lambda N/2)}{\sin(\lambda/2)} \right| ,$$

we observe that, for any $\eta > 0$, $\sup_{N \geq 1} \sup_{\lambda \in [\eta, 2\pi-\eta]} |D_N(\lambda)| < \infty$. Hence, with (61) and (63), we obtain Assumption 1 (iii).

Under (K-2) and (K-3), using that $K(t) = (2\pi)^{-1} \int \hat{K}(\xi) \, \mathrm{e}^{\mathrm{i}\xi t} \mathrm{d}\xi$, we get

$$\Phi(\lambda; i, v) = (2\pi \rho_{j-i,T})^{-1} \int_{-\infty}^{\infty} \hat{K}(\xi) \, \mathrm{e}^{\mathrm{i}\xi(uT_{j_i} - v)/(\mathrm{b}T_{j-i})} \sum_{l \in \tilde{\mathcal{T}}_j(i,v)} \mathrm{e}^{\mathrm{i}l(\lambda + 2^i\xi/(\mathrm{b}T_{j-i}))} \, \mathrm{d}\xi ,$$

where $\tilde{\mathcal{T}}_j(i,v)$ denotes the set of all $l \in \mathcal{T}_j(i,v)$ such that $\gamma_{j-i,T}(2^i l + v)$ does not vanish. Denote the length of $\tilde{\mathcal{T}}_j(i,v)$ by $N = N_{j,T}$ as in the previous case. We thus obtain

$$|\Phi(\lambda; i, v)| \leq (2\pi \rho_{j-i,T})^{-1} \int_{-\infty}^{\infty} \left| \hat{K}(\xi) \right| \, |D_N(\lambda + 2^i\xi/(\mathrm{b}T_{j-i}))| \, \mathrm{d}\xi .$$

Let $\eta > 0$. Splitting the above integral as $\int_{-\infty}^{\infty} = \int_{2^i|\xi|/(\mathrm{b}T_{j-i}) \leq \eta/2} + \int_{2^i|\xi|/(\mathrm{b}T_{j-i}) > \eta/2}$, we obtain

$$\begin{aligned} \sup_{\lambda \in [\eta, \pi]} |\Phi(\lambda; i, v)| \leq & (2\pi \rho_{j-i,T})^{-1} \sup_{|\lambda| \in [\eta/2, \pi+\eta/2]} |D_N(\lambda)| \\ & + (2\pi \rho_{j-i,T})^{-1} \|D_N\|_\infty \int_{2^i|\xi|/(\mathrm{b}T_{j-i}) > \eta/2} \left| \hat{K}(\xi) \right| \mathrm{d}\xi . \end{aligned}$$

Now, we have, for η small enough, $\sup_{N \geq 1} \sup_{|\lambda| \in [\eta/2, \pi+\eta/2]} |D_N|(\lambda) < \infty$, $\|D_N\|_\infty \leq N$ and, under (K-2), $N = O(\mathrm{b}T_j)$ and $\int_{2^i|\xi|/(\mathrm{b}T_{j-i}) > \eta/2} \left| \hat{K}(\xi) \right| \mathrm{d}\xi = o((\mathrm{b}T_j)^{-1/2})$, which, with the previous display, (61) and (63), implies Assumption 1 (iii). Under (K-3), the same conclusion holds using that $N = O(T_j)$, $\int_{2^i|\xi|/(\mathrm{b}T_{j-i}) > \eta/2} \left| \hat{K}(\xi) \right| \mathrm{d}\xi = O(\exp(-c 2^{-i-1} \eta \mathrm{b}T_j))$ and $T_j \exp(-c' \mathrm{b}T_{j-i}) = O(1)$ with $c' = c 2^{-i-1} \eta$.

Finally we show that Assumption 1 (iv) holds under (K-1), (K-2) and (K-3), successively. Using the definition (22) and (14), we get, for some positive constant C,

$$\Gamma_q(u;j,T) \le C\frac{(\mathrm{b}T_j)^q}{\rho_{j,T}}\sum_{k=0}^{T_j-1} K_q((uT_j-k)/(\mathrm{b}T_j)) + O(\Gamma_0(u;j,T)) ,$$

where $K_q(x)=K(x)|x|^q$. By definition of $\rho_{j,T}$, one has $\Gamma_0(u;j,T)=O(1)$. Under (K-1) and (K-2), K_q is bounded and compactly supported, so that $\sum_k K_q((uT_j-k)/(\mathrm{b}T_j)) = O(\mathrm{b}T_j)$. This, with (63) and the previous display, implies (30) for all $q\ge 0$. Hence, to conclude the proof, it only remains to show that, for $q=1,2$, under (K-3),

$$\sum_{k=0}^{T_j-1} K_q((uT_j-k)/(\mathrm{b}T_j)) = O(\mathrm{b}T_j) .$$

Using that $K(x)=O(|x|^{-p_0})$ as $x\to\pm\infty$, and $q\le 2$, we separate the sum $\sum_{k=0}^{T_j-1}$ in $\sum_{|uT_j-k|\le \mathrm{b}T_j}$ for which $K_q((uT_j-k)/(\mathrm{b}T_j))$ is $O(1)$ and $\sum_{|uT_j-k|>\mathrm{b}T_j}$ for which $K_q((uT_j-k)/(\mathrm{b}T_j))$ is $O(|(uT_j-k)/(\mathrm{b}T_j)|^{2-p_0})$. Hence, we get

$$\sum_{k=0}^{T_j-1} K_q((uT_j-k)/(\mathrm{b}T_j)) = O\left(\mathrm{b}T_j\right) + O\left(\mathrm{b}T_j^{p_0-2}\sum_{l\ge \mathrm{b}T_j-1} l^{2-p_0}\right) .$$

Observing that $\mathrm{b}T_j\to\infty$ and $p_0>3$, we obtain the desired bound. □

Lemma 5. *Suppose that* $\mathrm{b}_T\to 0$ *and* $T_j\mathrm{b}_T\to\infty$. *Then, for weights given by (34), Assumption 1 is satisfied with*

$$\delta_{j,T}\sim(\mathrm{b}_TT_j)^{-1} \tag{67}$$

$$V(i,v;i',v')=\pi\,2^{-i-i'},\quad i,i'\ge 0,\ v\in\{0,\dots,2^i-1\},\ v'\in\{0,\dots,2^{i'}-1\} . \tag{68}$$

Proof. For convenience, we will omit the subscripts $_T$ and $_{j,T}$ in this proof when no ambiguity arises. We set $u_j=[u\,T_j]$ in the following. Using (33), $\mathrm{b}T_j\to\infty$, $\mathrm{b}\to 0$ and $u_j\sim uT_j$, we get that

$$\rho\sim(\mathrm{b}T_j) . \tag{69}$$

Observing that $\delta=\gamma(u_j)=\rho^{-1}$, we get (67) and Assumption 1(i) follows.

Let us now show that Assumption 1(ii) holds. Using (64), we find that

$$\begin{aligned}\int_{-\pi}^{\pi}\Phi_{j,T}(\lambda;i,v)\overline{\Phi_{j,T}(\lambda;i',v')}\,\mathrm{d}\lambda =& \frac{2\pi}{\rho_{j-i,T}\rho_{j-i',T}}\exp\left(-\frac{u_{j-i}+v+1}{\mathrm{b}T_{j-i}}-\frac{u_{j-i'}+v'+1}{\mathrm{b}T_{j-i'}}\right)\\ &\times\sum_{l=0}^{N-1}\mathrm{e}^{l\{2^i/(\mathrm{b}T_{j-i})+2^{i'}/(\mathrm{b}T_{j-i'})\}} ,\end{aligned}$$

where $N=\{2^{-i}(u_{j-i}-v)\}\wedge\{2^{-i'}(u_{j-i'}-v')\}$. Using (14), (69), $\mathrm{b}T_j\to\infty$, $\mathrm{b}\to 0$ and $u_j\sim uT_j$, we obtain $\rho_{j-i,T}\sim 2^i(\mathrm{b}T_j)$, $(u_{j-i}+v+1)/(\mathrm{b}T_{j-i})\sim u/\mathrm{b}$, $2^i/(\mathrm{b}T_{j-i})\sim 1/(\mathrm{b}T_j)$, $N2^i/(\mathrm{b}T_{j-i})\sim u/\mathrm{b}$ and similar result with i',v' replacing i,v. Using these asymptotic equivalences and the previous display, we obtain

$$\int_{-\pi}^{\pi}\Phi_{j,T}(\lambda;i,v)\overline{\Phi_{j,T}(\lambda;i',v')}\,\mathrm{d}\lambda\sim\frac{2\pi}{2^{i+i'}(\mathrm{b}T_j)^2}\frac{A-o(1)}{2/(\mathrm{b}T_j)} , \tag{70}$$

where

$$A = \exp\left(-\frac{u_{j-i}+v+1}{\mathrm{b}T_{j-i}} - \frac{u_{j-i'}+v'+1}{\mathrm{b}T_{j-i'}} + N\left\{\frac{2^i}{\mathrm{b}T_{j-i}} + \frac{2^{i'}}{\mathrm{b}T_{j-i'}}\right\}\right) .$$

Using (14), we have $N = uT_j + O(1)$ and $u_{j-i} + v + 1 = uT_j 2^i + O(1)$. Thus $N2^i - (u_{j-i} + v + 1) = O(1)$ and the same holds with i', v' replacing i, v. This implies that $A = \exp\left(O\left((\mathrm{b}T_j)^{-1}\right)\right) \to 1$. This, (70) and (67) yield Assumption 1(ii) with $V(i, v; i', v')$ defined by (68).

We finally show that Assumption 1(iii) holds. By setting $N = 2^{-i}(u_{j-i} + v)$ and $k = N - 1 - l$ in (26), we obtain

$$|\Phi(\lambda; i, v)| = \rho^{-1}\left|\sum_{k=0}^{N-1} \mathrm{e}^{-k\{\mathrm{i}\lambda + 2^i/(\mathrm{b}T_{j-i})\}}\right| \leq \rho^{-1}\frac{1 + \mathrm{e}^{-N2^i/(\mathrm{b}T_{j-i})}}{\left|1 - \mathrm{e}^{-\mathrm{i}\lambda - 2^i/(\mathrm{b}T_{j-i})}\right|} .$$

Using that $N2^i/(\mathrm{b}T_{j-i}) \sim \mathrm{b}^{-1} \to \infty$, $\delta^{-1/2}\rho^{-1} \to 0$ and that, for any $\eta > 0$, $|1-z|$ does not vanish on the compact set of complex numbers $z = r\mathrm{e}^{\mathrm{i}\theta}$ such that $r \in [0, 1]$ and $\eta \leq |\theta| \leq \pi$ and thus is lower bounded on this set, we obtain Assumption 1(iii).

Finally we show that Assumption 1 (iv) holds. By (14), we have, for any $q \geq 0$,

$$\Gamma_q(u; j, T) = \rho^{-1}\sum_{k=0}^{u_j-1} \mathrm{e}^{-(u_j-1-k)/(\mathrm{b}T_j)}|u_j - 1 - k|^q + O\left(\Gamma_0(u; j, T)\right) .$$

Observe that $\Gamma_0(u; j, T) = 1$. Setting $l = u_j - 1 - k$, and separating the above sum over $l \leq [q\mathrm{b}T_j]+1$ for which we bound the exponential by 1 and $l \geq [q\mathrm{b}T_j]+2$ so that $\mathrm{e}^{-x/(\mathrm{b}T_j)}x^q$ is decreasing on $x \geq l - 1$, we get

$$\begin{aligned}\sum_{k=0}^{u_j-1} \mathrm{e}^{-(u_j-1-k)/(\mathrm{b}T_j)}|u_j - 1 - k|^q &\leq \sum_{l=0}^{[\mathrm{b}T_j]+1} l^q + \sum_{l\geq[\mathrm{b}T_j]+2} \mathrm{e}^{-l/(\mathrm{b}T_j)} l^q \\ &\leq O\left((\mathrm{b}T_j)^{q+1}\right) + \int_{x\geq[q\mathrm{b}T_j]+1} \mathrm{e}^{-x/(\mathrm{b}T_j)}x^q \mathrm{d}x \\ &= O\left((\mathrm{b}T_j)^{q+1}\right) .\end{aligned}$$

The last two displays, (69) and (67) yield (30), which achieves the proof. □

References

[1] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Inform. Theory*, 44(1):2–15, 1998. ISSN 0018-9448.

[2] J. Beran. On parameter estimation for locally stationary long-memory processes. Preprint, University of Konstanz, 2006.

[3] R. Dahlhaus. A likelihood approximation for locally stationary processes. *Ann. Statist.*, 28(6):1762–1794, 2000. ISSN 0090-5364.

[4] R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.*, 62(1):139–168, 1996. ISSN 0304-4149.

[5] R. Dahlhaus. Fitting time series models to non-stationary processes. *Annals of Statistics*, 25:1–37, 1997.

[6] Z. Ding, R. Engle, and C. Granger. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.

[7] G. Faÿ, F. Roueff, and P. Soulier. Estimation of the memory parameter of the infinite-source Poisson process. *Bernoulli*, 13(2):473–491, 2007. ISSN 1350-7265.
[8] G. Faÿ, E. Moulines, F. Roueff, and M.S. Taqqu. Estimators of long-memory: Fourier versus wavelets. *J. of Econometrics*, 2009. (In press, DOI:10.1016/j.jeconom.2009.03.005).
[9] D. Guégan and Z. Lu. Wavelet methods for locally stationary seasonal long memory processes. CES Working Paper 2009.15, Centre d'Economie de la Sorbonne, 2009.
[10] M. J. Jensen and B. Whitcher. Time-varying long memory in volatility: Detection and estimation with wavelets. Preprint, University of Missouri and Eurandom, 2000.
[11] E. Moulines, P. Priouret, and F. Roueff. On recursive estimation for time varying autoregressive processes. *Ann. Statist.*, 33(6):2610–2654, 2005.
[12] 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. *J. Time Ser. Anal.*, 28(2), 2007.
[13] E. Moulines, F. 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(4):301–313, 2007.
[14] E. Moulines, F. Roueff, and M.S. Taqqu. A wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series. *Ann. Statist.*, 36(4):1925–1956, 2008.
[15] P. M. Robinson. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23:1048–1072, 1995.
[16] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23:1630–1661, 1995.
[17] S.-H. Wang, C. Hsiao, and L. Bauwens. Forecasting long-memory processes subject to structural breaks. Preprint, CORE, Université catholique de Louvain, 2008.
[18] G. W. Wornell and A. V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Process.*, 40(3):611 – 623, March 1992.

Institut Telecom, Telecom Paris, CNRS LTCI, 46 rue Barrault, 75634 Paris Cedex 13, France

E-mail address: roueff@telecom-paristech.fr

Institut de statistique, Universit catholique de Louvain, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium

E-mail address: rvs@uclouvain.be