CARF Working Paper

CARF-F-527

Inference of Jumps Using Wavelet Variance

Heng Chen
Currency Department, The Bank of Canada

Mototsugu Shintani
Faculty of Economics, The University of Tokyo

November, 2021

CARF is presently supported by Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, The Dai-ichi Life Insurance Company, Limited, The Norinchukin Bank, MUFG Bank, Ltd. and Ernst & Young ShinNihon LLC. This financial support enables us to issue CARF Working Papers.

INFERENCE OF JUMPS USING WAVELET VARIANCE

BY HENG CHEN[1], MOTOTSUGU SHINTANI[2]

[1] *Currency Department, The Bank of Canada, chhe@bankofcanada.ca*

[2] *Faculty of Economics, The University of Tokyo, shintani@e.u-tokyo.ac.jp*

We consider the statistical inference of jumps in nonparametric regression models with long memory noise. A test statistic is proposed for the presence of jumps based on a robust estimator of the variance of the wavelet coefficients. The sequential applications of tests allow us to estimate the number of jumps and their locations. In comparison with the existing inference procedure, in which test statistic converges very slowly to the extreme value distribution, ours processes a more accurate finite sample performance derived from the asymptotic normality of our test statistic.

1. Introduction. We consider the model

$$Y_i = f(i/n) + \varepsilon_i, \tag{1}$$

for $i = 1, ..., n$, where $f(x)$ is an unknown nonparametric regression function defined on $[0, 1]$ and ε_i is a zero mean Gaussian error term allowing for long-range dependence. We observe the data $\{Y_1, ..., Y_n\}$ but the trend function $f(x)$ and noise ε_i are not separately observed. The objective of the analysis is two-fold. First, we consider the statistical inference on the presence of structural breaks in the form of abrupt mean shifts, or jumps, in the trend function $f(x)$. Second, once the presence of jumps is confirmed by the test, we estimate the number of jumps and their locations. To achieve these goals, we utilize the convenient properties of the wavelet transformation of the data. While the use of the wavelet in the detection of jumps has been also considered by Wang (1995, 1999), a notable feature of our proposed procedure is that the test statistic is constructed on the basis of the wavelet variance instead of the supremum of wavelet coefficients.

The nonparametric inference of regression functions with jumps in (1) has been an active area of research. In the context of the kernel estimation of $f(x)$, Müller (1992), Wu and Chu (1993), Qiu and Yandell (1998), Spokoiny (1998), Müller and Stadtmüller (1999), Gijbels and Goderniaux (2004), Gao, Gijbels and Bellegem (2008), and Porter and Yu (2015), among others, have investigated various procedures of jump detection (see also references therein for further information). Nevertheless, for most cases, the error term ε_i is assumed to be serially independent. Such a restriction is particularly problematic in time series analysis, in which dependence is the rule rather than the exception. The work of Wu and Zhao (2007) is one of the few exceptions to allow the serial dependence of ε_i in jump detection, but they consider only the case of short-range dependence. In contrast, our procedure, which is based on the wavelet variance estimation, allows for the long-range dependence of ε_i in (1).[1] The wavelet-based procedure of Wang (1995, 1999) also allows for long-range dependence.[2] However, under the null hypothesis of no jump, Wang's sup-type test statistic converges very slowly to

Keywords and phrases: additive outlier, jumps, long memory, nonparametric regression, wavelet transformation, wavelet variance.

[1] Under the parametric trend with long-memory noise, see Krämer and Sibbertsen (2002), Lazarová (2005) and Lavielle and Moulines (2000) for the inference of jumps.

[2] Wang (1995) focuses on ε_i to be independent Gaussian errors, while Wang (1999) studies ε_i as fractional Gaussian errors.

an extreme-value distribution, while our test statistic based on the wavelet variance converges faster to normal distribution. As a result, our procedure works better than Wang's procedure in the finite sample, which has been confirmed by simulation experiments.

Following the notations of Müller and Stadtmüller (1999), we now introduce some conditions on the trend function $f(x)$ and noise ε_i. The trend function $f(x)$ may be given by $f(x) \equiv f_C(x)$, where

$$f_C : [0,1] \to \mathbf{R},\ f_C \text{ is continuously differentiable, } \sup_{0\le x\le 1} \left|f_C'(x)\right| \le M.$$

Since the trend function is continuously differentiable, the number of jumps m_0 is 0, in this case. Alternatively, the trend function $f(x)$ may consist of the continuous part $f_C(x)$ and the discontinuous part $f_J(x)$, or $f(x) \equiv f_C(x) + f_J(x)$ where

$$f_J(x) \equiv \sum_{l=1}^{m_0} d_l I\{x \ge \lambda_l\},$$

for $x \in [0,1]$. Here, $m_0 \in \{1,2,3,...\}$ is the finite number of jumps, λ_l's are jump locations satisfying $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{m_0} < 1$, and d_l's are jump sizes with $d_l > 0$ for all $l \in \{1,..,m_0\}$. Regarding the condition on the noise ε_i, we assume that its autocorrelation of ε_i satisfies

$$corr\left(\varepsilon_i, \varepsilon_{i'}\right) \asymp \left|i - i'\right|^{-2+2H},\ i - i' \to \infty,$$

for $H \in [0.5, 1)$, where $a_n \asymp b_n$ if $0 < \liminf_{n\to\infty} |a_n/b_n| \le \limsup_{n\to\infty} |a_n/b_n| < \infty$. Here, H represents the Hurst parameter. The process ε_i is an independent Gaussian error when $H = 0.5$. The ε_i has long-range dependence (or long memory) when $H \in (0.5, 1)$.

The rest of the paper is organized as follows. In Section 2 we introduce two types of wavelet variance estimators, which play key roles in our test statistic. In Section 3, we propose a new procedure to detect jumps and to estimate the number of jumps. In Section 4 we revisit Wang's (1995, 1999) procedure and draw comparisons between Wang's and ours. A simulation experiment to evaluate the finite sample properties is carried out in Section 5. Section 6 contains applications of our procedure to the Dow-Jones Industrial Average and Nile river data. Proofs are given in Appendix A.

2. Wavelet Variance Estimation and Test Statistic.

2.1. *Discrete wavelet transformation.* We define $\mathbb{W}^{\mathbf{A}}_{j,k}$ as the discrete wavelet transformation (DWT) coefficient of $\{A_1, ..., A_n\}$ at scale $j \in Z$ and location $k \in Z$, such that,

$$\mathbb{W}^{\mathbf{A}}_{j,k} \equiv \frac{1}{n}\sum_{i=1}^{n} \psi_{j,k}\left(\frac{i}{n}\right) A_i$$

where

$$\psi_{j,k}\left(\frac{i}{n}\right) \equiv 2^{j/2}\psi\left(k - 2^j\frac{i}{n}\right),$$

and $\psi(t)$ is a wavelet function, that is, $\int \psi(t)\, dt = 0$. Mathematically, the wavelet coefficient is spatially adaptive to the pointwise smoothness of $f(x)$ (Donoho and Johnstone, 1995; Wang, 1995; Chen, Choi and Zhou, 2008), and the wavelet coefficient also decorrelates the long-memory noise ε_i (Wang, 1996); See Appendix A.3 for details. When there is no jump in the trend function $f(x)$, the wavelet coefficients are asymptotically Gaussian observations. On the other hand, when the trend function $f(x)$ has jumps, the magnitude of wavelet coefficients $\mathbb{W}^{\mathbf{Y}}_{j,k}$ at locations k near jumps will be large, and they essentially become outliers in the wavelet domain. Following spatial adaptivity and decorrelation properties, Wang (1995, 1999) proposed the test statistic $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ where $\mathbf{K}\equiv\left\{1,\cdots,2^{j}\right\}$ for jump detection because it would diverge in the presence of jumps.

Unlike Wang's approach, our test statistic is based on the second moment of wavelet coefficients. In particular, we utilize the wavelet variance, which measures the variability of wavelet coefficients (Percival, 1995). We define wavelet variance at a given scale j by

$$\sigma_j^2 \equiv Var\left(\mathbb{W}_{j,1}^{\mathbf{B}_H}\right)$$

where $\mathbb{W}_{j,k}^{\mathbf{B}_H}\equiv\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)\varepsilon_i$. Notice that $\sigma_j^2\asymp n^{2H-2}/2^{(2H-1)j}$, which is converging to 0 as $n\to\infty$.[3] There are two different estimators of wavelet variance σ_j^2 that have been frequently used in the literature. In what follows, we consider a new test statistic based on the difference of the two estimators of wavelet variance.

2.2. *Wavelet variance estimators.* The first estimator of wavelet variance is the so-called classical scale estimator, which has been considered in Serroukh, Walden and Percival (2000). It can be viewed as the sample variance in the wavelet domain, and is given by

$$\widehat{\sigma}_{j,\mathbf{K}}^2\equiv\frac{\sum_{k\in\mathbf{K}}(\mathbb{W}_{j,k}^{\mathbf{Y}})^2}{2^j},$$

for any j and $\mathbb{W}_{j,k}^{\mathbf{Y}}\equiv\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)Y_i$. Since the wavelet coefficients $\mathbb{W}_{j,k}^{\mathbf{Y}}$ are asymptotically Gaussian, $\widehat{\sigma}_{j,\mathbf{K}}^2$ is an estimator of σ_j^2 when there is no jump. On the other hand, when the trend function $f(x)$ contains a jump part $f_J(x)$, the wavelet coefficients $\mathbb{W}_{j,k}^{\mathbf{Y}}$ at locations k near jumps will become outliers in the wavelet domain. Since the presence of outliers typically leads to an inconsistency of sample variance, $\widehat{\sigma}_{j,\mathbf{K}}^2$ is not robust to the presence of jumps in $f(x)$.

The second estimator of wavelet variance is the square of median absolute deviation. In general, median absolute deviation is well-known for its robust property to outliers. In the context of the wavelet domain, it has been used in wavelet denoising applications (Donoho and Johnstone, 1994). Since the wavelet coefficients $\mathbb{W}_{j,k}^{\mathbf{Y}}$ are asymptotically centered Gaussian observations, the square of the median absolute deviation of $\left\{\mathbb{W}_{j,k}^{\mathbf{Y}}\right\}_{k=1}^{2^j}$ is defined as

$$\widetilde{\sigma}_{j,\mathbf{K}}^2\equiv\left[med_{k\in\mathbf{K}}\left|\frac{\mathbb{W}_{j,k}^{\mathbf{Y}}}{0.6745}\right|\right]^2.$$

Since median absolute deviation is robust to outliers, $\widetilde{\sigma}_{j,\mathbf{K}}^2$ is robust to the presence of jumps in $f(x)$.

2.3. *Test Statistic.* To construct a test statistics for the purpose of detecting jumps, we use the fact that $\widetilde{\sigma}_{j,\mathbf{K}}^2$ is a robust estimator of σ_j^2, regardless of the presence of jumps, and that

[3] Our parameter of interest σ_j^2 is shrinking to 0 and its convergence rate depends on H. This feature is similar to a feature in Kouamo et al. (2013), which they use to estimate the memory parameter. Also note that in Kouamo et al. (2013), their definition of wavelet variance instead diverges to infinity at the rate $2^{(2H-1)j}$, which is different from ours. This result occurs because: first, our $\psi_{j,k}(t)$ is defined as $2^{j/2}\psi\left(k-2^jt\right)$, while their $\psi_{j,k}(t)$ is $2^{-j/2}\psi\left(k-2^{-j}t\right)$; second, our wavelet transformation $\mathbb{W}_{j,k}^{\mathbf{A}}$ is based on the sample analogue of $\int\psi_{j,k}(t)A(dt)$, while Kouamo et al. (2013) simply use $\int\psi_{j,k}(t)A(dt)$.

$\widehat{\sigma}^2_{j,\mathbf{K}}$ is not robust to the presence of outliers in the wavelet domain. We define our (infeasible) test statistic $D_{j,\mathbf{K}}$ as

$$D_{j,\mathbf{K}} \equiv \frac{\widehat{\sigma}^2_{j,\mathbf{K}} - \widetilde{\sigma}^2_{j,\mathbf{K}}}{\sqrt{\omega}}$$

where

$$\omega \equiv (1,-1)\begin{pmatrix} v_1 & v_{12} \\ v_{12} & v_2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

and

$$v_1 \equiv Var\left[\frac{\sum_{k=1}^{2^j}\left(\mathbb{W}^{\mathbf{B}_H}_{j,k}\right)^2}{2^j}\right]$$

$$v_2 \equiv Var\left\{\left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{\mathbb{W}^{\mathbf{B}_H}_{j,k}}{0.6745}\right|\right]^2\right\}$$

$$v_{12} \equiv Cov\left\{\frac{\sum_{k=1}^{2^j}\left(\mathbb{W}^{\mathbf{B}_H}_{j,k}\right)^2}{2^j}, \left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{\mathbb{W}^{\mathbf{B}_H}_{j,k}}{0.6745}\right|\right]^2\right\}.$$

The intuition of why $D_{j,\mathbf{K}}$ can detect jumps is as follows. When the trend function $f(x)$ is smooth (i.e., $f_J(x) = 0$ and $m_0 = 0$), both $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$ are consistent estimators of the shrinking target σ^2_j, and the difference $\widehat{\sigma}^2_{j,\mathbf{K}} - \widetilde{\sigma}^2_{j,\mathbf{K}}$ should be small. On the other hand, when $f_J(x) \neq 0$ and $m_0 > 0$, the difference $\widehat{\sigma}^2_{j,\mathbf{K}} - \widetilde{\sigma}^2_{j,\mathbf{K}}$ would be large because $\widetilde{\sigma}^2_{j,\mathbf{K}}$ is the outlier-robust wavelet variance estimator, while $\widehat{\sigma}^2_{j,\mathbf{K}}$ is not a robust estimator in the presence of outliers in the wavelet domain. Moreover, note that the normalizer in the denominator of the test statistic $D_{j,\mathbf{K}}$ is not the squared root of $Var\left(\widehat{\sigma}^2_{j,\mathbf{K}} - \widetilde{\sigma}^2_{j,\mathbf{K}}\right) = Var\left(\frac{\sum_{k\in\mathbf{K}}(\mathbb{W}^{\mathbf{Y}}_{j,k})^2}{2^j} - \left[med_{k\in\mathbf{K}}\left|\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{0.6745}\right|\right]^2\right)$, which behaves differently under H_0 and H_1. Instead we use the squared root of $\omega \equiv Var\left(\frac{\sum_{k=1}^{2^j}\left(\mathbb{W}^{\mathbf{B}_H}_{j,k}\right)^2}{2^j} - \left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{\mathbb{W}^{\mathbf{B}_H}_{j,k}}{0.6745}\right|\right]^2\right)$, because such a normalizer does not depend on the presence of jumps.

3. Inference of Jumps. In this section, we first study the properties of the infeasible test statistic $D_{j,\mathbf{K}}$ in detecting jumps. Second, we estimate the number of jumps m_0 and their locations $\{\lambda_l\}_{l=1}^{m_0}$ based on $D_{j,\mathbf{K}}$. Third, we replace the denominator of $D_{j,\mathbf{K}}$ by a consistent standard error and propose a feasible test statistic $\widehat{D}_{j,\mathbf{K}}$.

3.1. *Testing for the hypothesis of no jumps.* We use the test statistic $D_{j,\mathbf{K}}$ for the purpose of the testing $H_0 : f_J(x) = 0$ and $m_0 = 0$ against $H_1 : f_J(x) \neq 0$ and $m_0 > 0$. To derive the asymptotic distribution of $D_{j,\mathbf{K}}$, we introduce the following set of the assumptions.

Assumption 1: $H - M \leq -1/2$, where M is the vanishing moment of ψ, i.e. $\int_{-\infty}^{\infty} t^m \psi(t)\, dt = 0$ for $m = 0, 1, ..., M-1$ and $\int_{-\infty}^{\infty} t^M \psi(t)\, dt \neq 0$.

Assumption 2: $\min_{1\leq l\leq m_0-1}(\lambda_{l+1} - \lambda_l) > \frac{T}{2^j} > 0$ and $\lambda_1, 1 - \lambda_{m_0} \geq \frac{T}{2^j} > 0$, where the support of wavelet function ψ is $[0, T]$ with $T < \infty$.

Assumption 3: As $n \to \infty$, we have (i) $\frac{n^{(H-1)/(H-3/2)}}{2^j} \to 0$; (ii) $\frac{n^{(2H-2)/(2H-5/2)}}{2^j} \to \infty$.

Assumption 1 imposes a restriction on the relationship between the Hurst parameter H and the vanishing moment M. This assumption implies that a sufficiently large vanishing moment M is required to achieve the decorrelation of the wavelet coefficients. Assumption 2 ensures that distances between jump locations will not be too short asymptotically, so that there is at most one jump in the interval of the length $T/2^j$. Assumption 3 imposes conditions on the divergence rate of the scale j. Assumption 3 (i) suggests that the scale j should not diverge too slowly to infinity to ensure the non-degenerating limiting distribution of $D_{j,\mathbf{K}}$ under H_0. At the same time, Assumption 3(ii) suggests that the scale j should not diverge too quickly to infinity, so that $D_{j,\mathbf{K}}$ diverges to infinity under H_1 and the test would be consistent. When $H = 0.5$ (the i.i.d. Gaussian error), Assumption 3 simplifies to $2^j \asymp n^\beta$ with $\beta \in (1/2, 2/3)$.

Let $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ be the probability and quantile function of the standard Gaussian variable, and $C_\gamma = \Phi^{-1}\left(1 - \frac{\gamma}{2}\right)$, where $\gamma \in (0,1)$. The following theorem provides asymptotic properties of $D_{j,\mathbf{K}}$ under H_0 and H_1.

THEOREM 3.1. *Suppose Assumptions 1,2 and 3 hold. Then*
(i) Under H_0 :

$$\lim_{n\to\infty} \Pr\left[|D_{j,\mathbf{K}}| \geq C_\gamma\right] = \gamma;$$

(ii) Under H_1 :

$$\lim_{n\to\infty} \Pr\left[|D_{j,\mathbf{K}}| \geq C_\gamma\right] = 1.$$

From the theorem above, we can test the presence of jumps by comparing $|D_{j,\mathbf{K}}|$ and the critical value C_γ at the $100 \times \gamma$ percent significance level from the quantile of the standard Gaussian distribution. When $|D_{j,\mathbf{K}}| \geq C_\gamma$, the null hypothesis H_0 is rejected against the alternative hypothesis H_1, which suggests the presence of jumps. As for the power of the test, the second part of the theorem suggests that our test is consistent under H_1.

For an intuitively understand of the first part of the theorem, namely, the asymptotic normality of the test statistic under H_0, it might be helpful to investigate separately the asymptotic behavior of $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$ with an appropriate normalizer. Since $\sigma_j^2 \asymp n^{2H-2}/2^{(2H-1)j}$, the limit of rescaled $\frac{2^{(2H-1)j}}{n^{2H-2}}\sigma_j^2$ is a positive constant. We can also apply the central limit theorem to the rescaled versions of $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$ to obtain

$$\begin{aligned}&\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widehat{\sigma}^2_{j,\mathbf{K}} - \sigma_j^2\right)\\ &= 2^{j/2}\left[\frac{2^{(2H-1)j}}{n^{2H-2}}\widehat{\sigma}^2_{j,\mathbf{K}} - \frac{2^{(2H-1)j}}{n^{2H-2}}\sigma_j^2\right] \xrightarrow{d} N\left(0,v_1^*\right)\end{aligned}$$

and

$$\begin{aligned}&\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widetilde{\sigma}^2_{j,\mathbf{K}} - \sigma_j^2\right)\\ &= 2^{j/2}\left[\frac{2^{(2H-1)j}}{n^{2H-2}}\widetilde{\sigma}^2_{j,\mathbf{K}} - \frac{2^{(2H-1)j}}{n^{2H-2}}\sigma_j^2\right] \xrightarrow{d} N\left(0,v_2^*\right)\end{aligned}$$

where

$$\begin{aligned}v_1^* &\equiv \lim_{n\to\infty} \frac{2^{(4H-1)j}}{n^{4H-4}} v_1\\ v_2^* &= \lim_{n\to\infty} \frac{2^{(4H-1)j}}{n^{4H-4}} v_2.\end{aligned}$$

By combining the limiting behaviors of $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$, we have

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}\right)\xrightarrow{d}N\left(0,\omega^*\right)$$

where

$$\omega^*\equiv\lim_{n\to\infty}\frac{2^{(4H-1)j}}{n^{4H-4}}\cdot\omega=(1,-1)\begin{pmatrix}v_1^* & v_{12}^*\\ v_{12}^* & v_2^*\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}$$

$$v_{12}^*\equiv\lim_{n\to\infty}\frac{2^{(4H-1)j}}{n^{4H-4}}v_{12}.$$

Therefore, under H_0, we have

$$D_{j,\mathbf{K}}=\frac{\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}\right)}{\sqrt{\frac{2^{(4H-1)j}}{n^{4H-4}}\cdot\omega}}\xrightarrow{d}N\left(0,1\right)$$

Next, for the second part of the theorem, namely, the asymptotic power of the test, we can understand the intuition by examining how the presence of jumps affects the asymptotic behavior of $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$, respectively. Under H_1, the asymptotic normality of $\widehat{\sigma}^2_{j,\mathbf{K}}$ can be obtained as

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)\xrightarrow{d}N\left(Bias_J,v_1^*\right)$$

where $Bias_J$ is the asymptotic bias caused by the presence of jumps. On the other hand, because of the robustness of $\widetilde{\sigma}^2_{j,\mathbf{K}}$, the asymptotic distribution of $\widetilde{\sigma}^2_{j,\mathbf{K}}$ under H_1 is given by

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widetilde{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)\xrightarrow{d}N\left(0,v_2^*\right),$$

which is the same as the limiting distribution under H_0. Thus, under H_1, we have

$$\begin{aligned}D_{j,\mathbf{K}}&=\frac{\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)-\left(\widetilde{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)}{\sqrt{\omega}}\\&=\frac{\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)-\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widetilde{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)}{\sqrt{\frac{2^{(4H-1)j}}{n^{4H-4}}\omega}}.\end{aligned}$$

Under Assumption 3(ii), $Bias_J\to\infty$, so that $\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)\to\infty$ as $n\to\infty$. By combining this result with $\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widetilde{\sigma}^2_{j,\mathbf{K}}-\sigma^2_j\right)=O_p\left(1\right)$ and $\sqrt{\frac{2^{(4H-1)j}}{n^{4H-4}}\omega}\to\omega^*<\infty$ as $n\to\infty$, we have $D_{j,\mathbf{K}}\to\infty$.

So far, the computation of $D_{j,\mathbf{K}}$ requires knowledge of ω. Later in Section 3.3, we replace ω with its estimator $\widehat{\omega}$ to obtain the feasible test statistic. It is important to note that our procedure does not require prior knowledge of the function $f_C(x)$ or of the Hurst parameter H. These advantages come from the following properties of wavelet transformation: (i) the wavelet coefficient of $f_C(x)$, $\mathbb{W}^{\mathbf{C}}_{j,k}$, is asymptotically of a smaller order than $\mathbb{W}^{\mathbf{J}}_{j,k}$ and $\mathbb{W}^{\mathbf{B}_H}_{j,k}$, so that its contribution to the test statistic is asymptotically negligible (Wang, 1995); (ii) the sequence of $\left\{\frac{2^{(H-1/2)j}}{n^{H-1}}\mathbb{W}^{\mathbf{B}_H}_{j,k}\right\}_{k=1}^{2^j}$ is asymptotically short-range dependent (Wang, 1996), so that the appearance of H is present only in the scaling parameter $\frac{2^{(H-1/2)j}}{n^{H-1}}$, which could be cancelled out from both the numerator and denominator of $D_{j,\mathbf{K}}$. Avoiding the estimation of the nonparametric trend function is desirable, because its convergence rate depends on the Hurst parameter (Hall and Hart, 1990). In addition, the precise estimation of the Hurst parameter is known to be difficult because it is sensitive to the choice of the tuning parameter (Faÿ et al., 2009).

3.2. *Estimating the number of jumps and their locations.* Here we consider the issue of estimating the number of jumps m_0 and their locations $\{\lambda_l\}_{l=1}^{m_0}$ by using the theoretical results from Section 3.1. Our approach relies on the sequential test using our test statistic $D_{j,\mathbf{K}}$, which is similar to those employed in the previous studies on the inference of multiple breaks, including Wang (1995) and Bai and Perron (1998). The procedure consists of multiple steps of computing the updated $D_{j,\mathbf{K}'}$ using a new set of wavelet coefficients indexed by $\mathbf{K}'$. Suppose we reject the null hypothesis of no jumps using $D_{j,\mathbf{K}}$ in the first step. In the second step, the neighborhood of the largest DWT coefficient $\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ included in the computation of $D_{j,\mathbf{K}}$ in the first step is removed because such DWT coefficients are likely to be outliers. This process is repeated until the latest $D_{j,\mathbf{K}'}$ fails to reject the null hypothesis of no additional jumps. The formal steps of this sequential procedure are described below.

Step 1 Conduct a test for $H_0 : m_0 = 0$ (no jump) against $H_1 : m_0 > 0$ (at least one jump). Reject H_0 if

$$|D_{j,\mathbf{K}}| > C_\gamma$$

where $\mathbf{K} \equiv \left\{1, \cdots, 2^j\right\}$. If H_0 is not rejected, set $\widehat{m} = 0$;

Step 2 If $H_0 : m_0 = 0$ is rejected in Step 1, conduct a test for $H_0 : m_0 = 1$ (one jump) against $H_1 : m_0 > 1$ (at least two jumps). Reject H_0 if

$$\left|D_{j,\mathbf{K}\backslash\widehat{\mathbf{K}}_1}\right| > C_\gamma$$

where

$$\widehat{\mathbf{K}}_1 \equiv \left\{k : \widehat{k}_1 - k \in \text{supp}(\psi)\right\}$$

with $\widehat{k}_1 \equiv \arg\sup_{k\in\mathbf{K}} \left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$. If H_0 is not rejected, set $\widehat{m} = 1$;

Step 3 If $H_0 : m_0 = 1$ is rejected, conduct a test for $H_0 : m_0 = 2$ (two jumps) against $H_1 : m_0 > 2$ (at least three jumps). Reject H_0 if

$$\left|D_{j,\mathbf{K}\backslash\left(\widehat{\mathbf{K}}_1\cup\widehat{\mathbf{K}}_2\right)}\right| > C_\gamma$$

where

$$\widehat{\mathbf{K}}_2 \equiv \left\{k : \widehat{k}_2 - k \in \text{supp}(\psi)\right\}$$

with $\widehat{k}_2 \equiv \arg\sup_{k\in\mathbf{K}\backslash\widehat{\mathbf{K}}_1} \left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$. If H_0 is not rejected, set $\widehat{m} = 2$;

Step 4 Repeat the step until H_0 is not rejected, so that $\widehat{m}$ satisfies

$$\left|D_{j,\mathbf{K}\backslash\cup_{l=1}^{\widehat{m}}\widehat{\mathbf{K}}_l}\right| \leq C_\gamma$$

where

$$\widehat{\mathbf{K}}_l \equiv \left\{k : \widehat{k}_l - k \in \text{supp}(\psi)\right\}$$

with $\widehat{k}_l \equiv \arg\sup_{k\in\mathbf{K}\backslash\cup_{l=1}^{l-1}\widehat{\mathbf{K}}_l} \left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ with $l = 1, \cdots, \widehat{m}$.

Following the above procedure, we can show that both the estimated number of jumps and the locations are consistent in the following theorem.

THEOREM 3.2. *Suppose Assumptions 1,2 and 3 hold and $\gamma \to 0$. Then*

$$\Pr(\widehat{m} = m_0) \to 1,$$

$$\sum_{l=1}^{\widehat{m}_0} \left(\widehat{\lambda}_l - \lambda_l\right)^2 = O_p\left(2^{-2j}\right)$$

where

$$\widehat{\lambda}_l \equiv \frac{\widehat{k}_l}{2^j}.$$

The first part of Theorem 2 shows that the number of jumps m_0 can consistently be estimated by letting $\gamma \to 0$ as $n \to \infty$. The second part of Theorem 2 shows that our procedure yields a consistent estimation of jump locations with the convergence rate 2^j. When the number of jumps m_0 is unknown, we can estimate jump locations with a convergence rate up to $n^{\frac{2H-2}{2H-5/2}} / (\log n)^{\kappa}$ with $\kappa > 0$. When $H = 0.5$, we have a faster convergence rate $n^{2/3} / (\log n)^{\kappa}$ with $\kappa > 0$ than the case where $H \in (0.5, 1)$.

3.3. *Feasible Test Statistic.* We now consider the estimation of ω in the denominator of $D_{j,\mathbf{K}}$ to obtain the feasible test statistic**.** Let us define

$$\widehat{U}_k \equiv \begin{pmatrix} \widehat{U}_{k1} \\ \widehat{U}_{k2} \end{pmatrix}$$

where

$$\widehat{U}_{k1} \equiv (\mathbb{W}^{\mathbf{Y}}_{j,k})^2 \cdot I\left\{\left|\mathbb{W}^{\mathbf{Y}}_{j,k}\right| \leq \widehat{Q}_{\left|\mathbb{W}^{\mathbf{Y}}_j\right|}(1-\epsilon)\right\} - \widetilde{\sigma}^2_{j,\mathbf{K}} \text{ and}$$

$$\widehat{U}_{k2} \equiv \frac{-1.4826 \cdot \widetilde{\sigma}^2_{j,\mathbf{K}}}{\varphi(0.6745)} \left[I\left\{\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{\widetilde{\sigma}_{j,\mathbf{K}}} \leq 0.6745\right\} - I\left\{\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{\widetilde{\sigma}_{j,\mathbf{K}}} \leq -0.6745\right\} - 0.5\right]$$
$$\times I\left\{\left|\mathbb{W}^{\mathbf{Y}}_{j,k}\right| \leq \widehat{Q}_{\left|\mathbb{W}^{\mathbf{Y}}_j\right|}(1-\epsilon)\right\}$$

with $\widehat{Q}_{\mathbb{W}^{\mathbf{Y}}_j}(1-\epsilon)$ being the empirical quantile of $\left|\mathbb{W}^{\mathbf{Y}}_{j,k}\right|$ at $(1-\epsilon)$, and $\varphi(\cdot)$ being the density of the standard normal distribution.

Then our proposed estimator of ω is given by

$$\widehat{\omega} \equiv (1, -1)\,\widehat{\Omega}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

where

$$\widehat{\Omega} \equiv \sum_{l=-h}^{h} \left(1 - \frac{|l|}{h+1}\right) \widehat{\Gamma}(l)$$

$$\widehat{\Gamma}(l) \equiv \frac{1}{2^j \cdot (1-\epsilon)} \sum_{1 \leq k-l \leq 2^j \cdot (1-\epsilon)} \widehat{U}_{k-l} \widehat{U}'_k.$$

Our proposed estimator of ω can be understood in three steps. In the first step, we rewrite the estimation errors of $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$ in terms of sample averages, which are

$$\widehat{\sigma}^2_{j,\mathbf{K}} - \sigma^2_j = \frac{\sum_{k\in\mathbf{K}}(\mathbb{W}^{\mathbf{Y}}_{j,k})^2}{2^j} - \frac{2^j\sigma^2_j}{2^j} = \frac{\sum_{k\in\mathbf{K}}\left[(\mathbb{W}^{\mathbf{Y}}_{j,k})^2 - \sigma^2_j\right]}{2^j} \tag{2}$$

$$\widetilde{\sigma}^2_{j,\mathbf{K}} - \sigma^2_j = \frac{-1.4826\cdot\sigma^2_j}{\varphi(0.6745)\cdot 2^j}\sum_{k\in\mathbf{K}}\left[I\left\{\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{\sigma_j}\le 0.6745\right\} - I\left\{\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{\sigma_j}\le -0.6745\right\} - 0.5\right]$$
$$+o_p(1). \tag{3}$$

Note that we used the asymptotic expansion result for $\widetilde{\sigma}^2_{j,\mathbf{K}}$ in Proposition 1 of Kouamo et al. (2013) to derive (3). In the second step, since the definition of ω is based on $\mathbb{W}^{\mathbf{B}_H}_{j,k}$ instead of $\mathbb{W}^{\mathbf{Y}}_{j,k}$, we truncate the $100\times(1-\epsilon)$ percent of the largest $\left|\mathbb{W}^{\mathbf{Y}}_{j,k}\right|$ to construct the truncated version of (2) for $\widehat{\sigma}^2_{j,\mathbf{K}}$, which is

$$\sum_{k\in\mathbf{K}}\frac{(\mathbb{W}^{\mathbf{Y}}_{j,k})^2\cdot I\left\{\left|\mathbb{W}^{\mathbf{Y}}_{j,k}\right|\le \widehat{Q}_{|\mathbb{W}^{\mathbf{Y}}_j|}(1-\epsilon)\right\} - \sigma^2_j}{2^j\cdot(1-\epsilon)}; \tag{4}$$

and the truncated version of (3) for $\widetilde{\sigma}^2_{j,\mathbf{K}}$ is

$$\sum_{k\in\mathbf{K}}\left[I\left\{\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{\sigma_j}\le 0.6745\right\} - I\left\{\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{\sigma_j}\le -0.6745\right\} - 0.5\right]\cdot I\left\{\left|\mathbb{W}^{\mathbf{Y}}_{j,k}\right|\le \widehat{Q}_{|\mathbb{W}^{\mathbf{Y}}_j|}(1-\epsilon)\right\}$$
$$\frac{-1.4826\cdot\sigma^2_j}{\varphi(0.6745)\cdot 2^j\cdot(1-\epsilon)}. \tag{5}$$

In the third step, we substitute σ^2_j with $\widetilde{\sigma}^2_{j,\mathbf{K}}$ in (4) and (5) to obtain

$$\frac{1}{2^j\cdot(1-\epsilon)}\sum_{k\in\mathbf{K}}\widehat{U}_{k1}$$
$$\frac{1}{2^j\cdot(1-\epsilon)}\sum_{k\in\mathbf{K}}\widehat{U}_{k2}.$$

Recall that $\omega \equiv Var\left(\frac{\sum_{k=1}^{2^j}\left(\mathbb{W}^{\mathbf{B}_H}_{j,k}\right)^2}{2^j} - \left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{\mathbb{W}^{\mathbf{B}_H}_{j,k}}{0.6745}\right|\right]^2\right)$ is defined according to $\mathbb{W}^{\mathbf{B}_H}_{j,k}$. Because the truncation is designed to remove the components of $\mathbb{W}^{\mathbf{Y}}_{j,k}$ that are related to jumps from asymptotic expansions, the remaining terms of $\mathbb{W}^{\mathbf{Y}}_{j,k}$ are asymptotically equivalent to $\mathbb{W}^{\mathbf{B}_H}_{j,k}$. Therefore, both $\widehat{U}_{k1}$ and $\widehat{U}_{k2}$ are not affected by the presence of jumps: $\widehat{U}_{k1}$ is asymptotically equivalent to $(\mathbb{W}^{\mathbf{B}_H}_{j,k})^2 - \sigma^2_j$ under both H_0 and H_1, and $\widehat{U}_{k2}$ is asymptotically equivalent to $\frac{-1.4826\cdot\sigma^2_j}{\varphi(0.6745)}\left[I\left\{\frac{\mathbb{W}^{\mathbf{B}_H}_{j,k}}{\sigma_j}\le 0.6745\right\} - I\left\{\frac{\mathbb{W}^{\mathbf{B}_H}_{j,k}}{\sigma_j}\le -0.6745\right\} - 0.5\right]$ under both H_0 and H_1. In order to prove the consistency of $\widehat{\omega}$ under both H_0 and H_1, we introduce an additional assumption as follows.

Assumption 4: (i) $\epsilon \ge 2^{-j}(Tm_{\max}+4)$; (ii) as $n\to\infty$, $h\to\infty$ and $h^3/2^j = O(1)$.

With Assumption 4(i), we need to choose ϵ such that $\epsilon \ge 2^{-j}(Tm_{\max}+4)$ where T is the length of the support of ψ and $m_{\max}$ is the maximum number of jumps. The reason for this lower bound of ϵ is that the number of truncated observations should be larger than the total number of wavelet coefficients affected by either jumps or boundaries near 0 or 1 (Percival and Walden, 2000); otherwise, our estimate $\widehat{\omega}$ will be affected by atypical observations in the wavelet domain. In addition, by removing these outliers, the automatic bandwidth selection of Andrews (1991) is valid. In practice, we prefer to choose $\epsilon \to 0$ as $n \to \infty$ so that we can have a precise $\widehat{\omega}$ by using as many wavelet as possible that are not affected by jumps.

Assumption 4(ii) requires that the bandwidth (or lag truncation number) h grows at a rate no faster than $2^{j/3}$. In what follows, we use the automatic bandwidth selection procedure of Andrews (1991) with the Bartlett kernel to choose h, which satisfies Assumption 4(ii).

Suppose Assumptions 1, 2, 3 and 4 hold. Then under both H_0 and H_1,

$$\frac{2^{(4H-1)j}}{n^{4H-4}}\left(\widehat{\omega}-\omega\right)=o_p\left(1\right).$$

The proposition implies that $\frac{2^{(4H-1)j}}{n^{4H-4}} \cdot \widehat{\omega}$ is a consistent estimator of $\omega^* \equiv \lim_{n\to\infty} \frac{2^{(4H-1)j}}{n^{4H-4}} \cdot \omega$. Because of Proposition 3, our Theorem 3.1 still holds by Slutsky's theorem if $\frac{2^{(4H-1)j}}{n^{4H-4}} \cdot \omega$ in

$$\begin{aligned} D_{j,\mathbf{K}} &\equiv \frac{\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}}{\sqrt{\omega}} \\ &= \frac{\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}\right)}{\sqrt{\frac{2^{(4H-1)j}}{n^{4H-4}}\omega}} \end{aligned}$$

is replaced by $\frac{2^{(4H-1)j}}{n^{4H-4}} \cdot \widehat{\omega}$.[4] Then our feasible test statistic is given by

$$\widehat{D}_{j,\mathbf{K}} \equiv \frac{\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}}{\sqrt{\widehat{\omega}}},$$

which replaces $D_{j,\mathbf{K}}$ in both the testing in Section 3.1 and the estimation of m_0 and λ_l in Sections 3.2.

Notice that it is important to truncate the $100 \times (1-\epsilon)$ percent of the largest $\left|\mathbb{W}^{\mathbf{Y}}_{j,k}\right|$ from the asymptotic expansions of $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$ (Equations 2 and 3); otherwise, the sample autocovariance will be influenced by outliers in the presence of jumps.[5] Such outliers will lead

[4] In the simulation (which is not reported), we also apply the resampling method of Carlstein (1986) to estimate ω with a robustified variance estimator developed by Rousseeuw and Croux (1993). More specifically, divide $\left\{\mathbb{W}^{\mathbf{Y}}_{j,k}\right\}_{k=1}^{2^j}$ into adjacent nonoverlapping subseries of length $m \equiv m_j$ with $m_j \to \infty$ and $m_j/2^j \to 0$, and compute $\left\{\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}\right)^l_m\right\}_{l=1}^{\lfloor 2^j/m \rfloor}$, where $\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}\right)_m$ is the estimate of $\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}$ computed from the subseries values of $\left(\mathbb{W}^{\mathbf{Y}}_{j,k+1},\ldots,\mathbb{W}^{\mathbf{Y}}_{j,k+m}\right)$. To ensure that the resampling estimated variance is robust to outliers, we propose

$$\begin{aligned} &\widetilde{\omega} \\ &\equiv \left(2.219\left\{\left|\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}\right)^l_m-\left(\widehat{\sigma}^2_{j,\mathbf{K}}-\widetilde{\sigma}^2_{j,\mathbf{K}}\right)^{l'}_m\right|; l<l'\right\}_{(s)}\right)^2 / \left\lfloor 2^j/m \right\rfloor \end{aligned}$$

where $s = \binom{\lfloor 2^j/m \rfloor}{2}/4$ and $\{\cdot\}_{(s)}$ is the s-th order statistic of $\binom{\lfloor 2^j/m \rfloor}{2}$ interpoint distances. For the choice of m (length of subseries), we follow Carlstein's (1986) optimal length computation based on an autoregressive AR(1) model with iid Gaussian innovations; see Section 5 of Carlstein (1986) for details. However, from our simulation, this resampling alternative $\widetilde{\omega}$ has a larger size distortion than $\widehat{\omega}$.

[5] See Section 3.1 from Qu and Perron (2010) for an example of the impacts of mean shifts on sample autocovariance.

to an upward bias in the estimation of an autocovariances of $\left\{\frac{2^{(2H-1/2)j}}{n^{2H-2}}\left(\mathbb{W}_{j,k}^{\mathbf{B}_H}\right)^2\right\}_{k=1}^{2^j}$, so that $\frac{2^{(4H-1)j}}{n^{4H-4}}\cdot\widehat{\omega}$ with $\epsilon=0$ will be an inconsistent estimator of ω^*.

4. Revisiting Wang's Test. Our inference of jumps is based on $\widehat{D}_{j,\mathbf{K}}$, in which the limiting distribution is normal under H_0. Recall that 2^j is the number of wavelet coefficients at scale j, such that $2^j/n\to 0$, and that then the empirical process theory suggests that the rate of convergence for $\widehat{D}_{j,\mathbf{K}}$ to asymptotic normal distribution is $\left(2^j/\log\log 2^j\right)^{1/2}$ (Van der Vaart, 2000, page 268). In contrast, the rate of convergence for $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ is extremely slow and very large values of n are needed for the approximation to be reasonably accurate. When we apply the generic results of Hall (1979) and Rootzén (1983), we see that when $H=0.5$, the rate convergence of $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ is $\log 2^j$; while for $H\in(0.5,1)$, the rate is proportional to $\log 2^j/\left(\log\log 2^j\right)^2$. Therefore, our inferential procedure has better size control than Wang's (1995, 1999) procedure, due to the convergence rate of $\widehat{D}_{j,\mathbf{K}}$ (to the asymptotic normal distribution), which is faster than the convergence rate of $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ (to the asymptotic extreme-value type distribution). Our simulation in the next section confirms that $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ tends to have a larger Type I error than the nominal size under H_0, and such a size distortion in $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ is much larger than that in $\widehat{D}_{j,\mathbf{K}}$.[6]

We first revisit the asymptotic property of Wang's (1995, 1999) test statistic $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ under H_0. Under some regularity conditions,

$$\lim_{n\to\infty}\Pr\left[\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|\geq C_\gamma^*\right]=\gamma$$

where C_γ^* is the estimated critical value from the *extreme-value* distribution. In particular, the critical value is given by

(6)
$$C_\gamma^*\equiv\widetilde{\sigma}_{j,\mathbf{K}}\cdot\left(\left\{2\left|\log\left(2^{-j}\right)\right|\right\}^{1/2}-\left\{2\left|\log\left(2^{-j}\right)\right|\right\}^{-1/2}\log\left\{-2^{1/2}\pi\cdot\widetilde{\sigma}_{j,\mathbf{K}}\cdot\widetilde{\delta}_{j,\mathbf{K}}\cdot\log\left(1-\gamma\right)\right\}\right)$$

where

$$\widetilde{\delta}_{j,\mathbf{K}}\equiv\left[med_{k\in\mathbf{K}}\left|\frac{\mathbb{W}_{j,k}^{(1)\mathbf{Y}}}{0.6745}\right|\right]^2,$$

and

$$\mathbb{W}_{j,k}^{(1)\mathbf{Y}}\equiv\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}^{(1)}\left(\frac{i}{n}\right)Y_i$$

and $\psi^{(1)}$ is the first derivative of ψ. In Equation (6) $\widetilde{\sigma}_{j,\mathbf{K}}$ and $\widetilde{\delta}_{j,\mathbf{K}}$ are empirical counterparts of τ_1 and τ_2 in Theorem 1 of Wang (1999). It should be noted that using τ_1 and τ_2 requires the knowledge of H and integral calculations that involve wavelet functions. Replacing τ_1 and

[6] There are other simulation studies documenting the poor finite sample performance of test statistic, which converges to the extreme-value distribution. For example, Wu and Chu (1993, Figure 4) show that the test statistic based on the supremum of the absolute value of kernel estimates has a large Type I error relative to the nominal size under H_0.

τ_2 by $\widetilde{\sigma}_{j,\mathbf{K}}$ and $\widetilde{\delta}_{j,\mathbf{K}}$ is convenient because the estimation of H is not needed. The above test statistic unifies Wang (1995) and Wang (1999) to allow for both independent and fractional Gaussian noises.[7]

In what follows, we describe the sequential procedure to estimate the number of jumps and their locations based on Wang's test statistic. Such a description helps us clarify the similarities and differences between our and Wang's procedure.

Step 1 Conduct a test for $H_0 : m_0 = 0$ (no jump) against $H_1 : m_0 > 0$ (at least one jump). Reject H_0 if

$$\sup_{k \in \mathbf{K}} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right| > C^*_\gamma.$$

If H_0 is not rejected, set $\widehat{m}^* = 0$;

Step 2 If $H_0 : m_0 = 0$ is rejected in Step 1, conduct a test for $H_0 : m_0 = 1$ (one jump) against $H_1 : m_0 > 1$ (at least two jumps). Reject H_0 if

$$\sup_{k \in \mathbf{K} \backslash \widehat{\mathbf{K}}^*_1} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right| > C^*_\gamma$$

where

$$\widehat{\mathbf{K}}^*_1 \equiv \left\{ k : \widehat{k}^*_1 - k \in \text{supp}(\psi) \right\}$$

with $\widehat{k}^*_1 \equiv \arg\sup_{k \in \mathbf{K}} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right|$. If H_0 is not rejected, set $\widehat{m}^* = 1$;

Step 3 If $H_0 : m_0 = 1$ is rejected, conduct a test for $H_0 : m_0 = 2$ (two jumps) against $H_1 : m_0 > 2$ (at least three jumps). Reject H_0 if

$$\sup_{k \in \mathbf{K} \backslash \left(\widehat{\mathbf{K}}^*_1 \cup \widehat{\mathbf{K}}^*_2 \right)} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right| > C^*_\gamma$$

where

$$\widehat{\mathbf{K}}^*_2 \equiv \left\{ k : \widehat{k}^*_2 - k \in \text{supp}(\psi) \right\}$$

with $\widehat{k}^*_2 \equiv \arg\sup_{k \in \mathbf{K} \backslash \widehat{\mathbf{K}}^*_1} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right|$. If H_0 is not rejected, set $\widehat{m}^* = 2$;

Step 4 Repeat the step until H_0 is not rejected, so that $\widehat{m}^*$ satisfies

$$\sup_{k \in \mathbf{K} \backslash \cup_{l=1}^{\widehat{m}^*} \widehat{\mathbf{K}}^*_l} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right| \le C^*_\gamma$$

where

$$\widehat{\mathbf{K}}_l \equiv \left\{ k : \widehat{k}^*_l - k \in \text{supp}(\psi) \right\}$$

with $\widehat{k}^*_l \equiv \arg\sup_{k \in \mathbf{K} \backslash \cup_{l=1}^{l-1} \widehat{\mathbf{K}}^*_l} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right|$ with $l = 1, \cdots, \widehat{m}^*$.

While the expressions of $\widehat{k}_l \equiv \arg\sup_{k \in \mathbf{K} \backslash \cup_{l=1}^{l-1} \widehat{\mathbf{K}}_l} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right|$ and $\widehat{k}^*_l \equiv \arg\sup_{k \in \mathbf{K} \backslash \cup_{l=1}^{l-1} \widehat{\mathbf{K}}^*_l} \left| \mathbb{W}^{\mathbf{Y}}_{j,k} \right|$ are the same between the two sequential procedures, there are two potential sources of the

[7] Although both ours and Wang's (1995, 1999) are consistent under H_1, their local alternatives are different. When the jump size $d_l = O\left(n^{-1/4}\right)$, the power of our test goes to 1. In contrast, the power of Wang (1995, 1999) goes to 1 when the jump size $d_l = O\left(n^{\frac{1-H}{H-2}}\right)$. Thus, if $H \in [0.5, 2/3]$ and the jump sizes are small, Wang's test is more powerful than ours in terms of the local alternative.

difference in results between $\left\{\widehat{k}_l\right\}_{l=1}^{\widehat{m}}$ and $\left\{\widehat{k}_l^*\right\}_{l=1}^{\widehat{m}^*}$. First, $\widehat{m}$ and $\widehat{m}^*$ can be different because our test, which is based on $D_{j,\mathbf{K}'}$, and Wang's test $\sup_{k\in\mathbf{K}'}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ can yield different results regarding the rejection of the null hypothesis of no additional jumps. Second, even when $\widehat{m}$ and $\widehat{m}^*$ are the same, $\widehat{k}_l$ and $\widehat{k}_l^*$ can be different if the choice 2^j differs between the two procedures. In fact, the required conditions on 2^j differs between the two procedures. In particular, Assumption 3 can be relaxed in Wang's procedure as follows:

Assumption 3*: As $n\to\infty$, we have (i) $\frac{n^{(H-1)/(H-2)}}{2^j}\to 0$; (ii) $\frac{n/\log n}{2^j}\to\infty$.

For example, when $H=0.5$ (the i.i.d. Gaussian error), Assumption 3* simplifies to $n^{1/3}/2^j\to 0$ and $n\left(\log n\right)^{-1}/2^j\to\infty$, that is, $2^j\asymp\zeta_n$ with $\zeta_n\in\left(n^{1/3},n/\log n\right)$. Note that unlike Assumption 3, the upper bound of 2^j in Assumption 3* (ii) does not depend on H.

(Wang 1995 and 1999) Suppose Assumptions 1,2 and 3* hold and $\gamma\to 0$. Then

$$\Pr\left(\widehat{m}^*=m_0\right)\to 1,$$

$$\sum_{l=1}^{\widehat{m}^*}\left(\widehat{\lambda}_l^*-\lambda_l\right)^2=O_p\left(2^{-2j}\right)$$

where

$$\widehat{\lambda}_l^*\equiv\frac{\widehat{k}_l^*}{2^j}.$$

The first part of the proposition shows that $\widehat{m}^*$ is consistent for the number of jumps m_0. This property is common between $\widehat{m}^*$ and $\widehat{m}$. However, the finite sample property of $\widehat{m}^*$ is not expected to be as good as $\widehat{m}$ because of the following reasons. Recall that the rate of convergence for $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$ is extremely slow and a very large sample size is needed for the approximation to be reasonably accurate. While a poor approximation in the finite sample can result in size distortion in either direction, Wu and Chu (1993) have documented that similar test statistic based on the supremum of kernel estimates of potential jump sizes had larger type I error than the nominal level. From the simulation in the next section, we also find that over-rejection of Wang's test in the finite sample.

In contrast, our test statistic $\widehat{D}_{j,\mathbf{K}}$ converges to the normal distribution with a rate faster than $\sup_{k\in\mathbf{K}}\left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$. In the following simulation section, we show that the Type I error of our test is quite close to the nominal level under the null hypothesis. Since the larger Type I error of Wang's test implies that the null hypothesis would be rejected more often than in our test, $\widehat{m}^*$ leads to a positive bias compared to $\widehat{m}$. The following simulation section confirms this advantage of $\widehat{m}$ over $\widehat{m}^*$.

The second part of the proposition shows that Wang's procedure yields a consistent estimation of jump locations with the convergence rate 2^j, which is exactly the same as in our procedure, which is shown in Theorem 2. However, it should be noted that allowable ranges of the divergence rate of 2^j differ between Theorem 2 and Proposition 4. For the convergence rates of two jump location estimators $\widehat{\lambda}_l$ and $\widehat{\lambda}_l^*$, we can compare their fastest rates by choosing 2^j, which satisfies Assumption 3 and 3*, respectively. For example, if we set $2^j\asymp n^{\frac{2H-2}{2H-5/2}}/\left(\log n\right)^{\kappa}$ which is allowed in Assumption 3, our jump location estimator $\widehat{\lambda}_l$ converges at a rate of $n^{\frac{2H-2}{2H-5/2}}/\left(\log n\right)^{\kappa}$ with $\kappa>0$. In contrast, if we set $2^j\asymp n/\left(\log n\right)^{\eta}$ where $\eta>1$, which is the fastest rate allowed in Assumption 3*, $\widehat{\lambda}_l^*$ converges at a rate of $n/\left(\log n\right)^{\eta}$, which is independent of H. For this reason, if we aim only for the convergence rate of the jump location, $\widehat{\lambda}_l^*$. dominates $\widehat{\lambda}_l$.[8] However, choosing the largest 2^j has potential drawbacks. For example, if ε_i is non-Gaussian, our assumption on the Gaussianity of $\mathbb{W}_{j,k}^{\mathbf{Y}}$ needs to be replaced by an approximation that crucially depends on the sample size $n/2^j$, which would be small when 2^j is too large. Furthermore, if we are interested in constructing confidence bands of the estimated smooth function, the optimal choice of 2^j would be then based on the mean-squared error.

In the following simulation section, we examine the relative performance of $\widehat{m}$ and $\widehat{m}^*$ based on the common scale and the different scale. While deriving the optimal rate is beyond the scope of this paper, we provide some guidance on choosing the scale in practice, based on the simulations. Because $\gamma \to 0$ is a condition in Proposition 4, Wang (1995, 1999) recommended the use of $\lim_{\gamma \to 0} C_\gamma^* = \left(2 \log 2^j\right)^{1/2} \widetilde{\sigma}_{j,\mathbf{K}}$ for the critical value. In the simulation, we also employ the critical values C_γ and C_γ^* by setting $\gamma = 0.05$ and 0.01, which could be more or less reasonable in the finite sample, and compare the performance of $\widehat{m}$ and $\widehat{m}^*$.

5. Simulation Study. In this section, we conduct a simulation experiment to evaluate the finite sample performance of our proposed method in Section 3 in comparison with that of Wang's procedure in Section 4. We choose either the smooth trend function $f_C(x) = 1$ (Wang, 1995) or $f_C(x) = 2\left[\sin(4\pi x) + \sin(8\pi x)\right]$ (Perron, Shintani and Yabu, 2020), and the jump part $f_J(x) = \sum_{l=1}^{m_0} d_l I\{x > \lambda_l\}$ with either $m_0 = 0$ or $m_0 = 3$. Therefore, we have four DGPs for the trend function $f(x)$:

DGP1 No jumps: $m_0 = 0$ and the smooth trend function $f_C(x) = 1$;
DGP2 No jumps: $m_0 = 0$ and the smooth trend function $f_C(x) = 2\left[\sin(4\pi x) + \sin(8\pi x)\right]$;
DGP3 Three jumps: $m_0 = 3$ with $d_1 = 2$ and $\lambda_1 = 0.25, d_2 = -2$ and $\lambda_2 = 0.50$, $d_3 = 2$ and $\lambda_3 = 0.75$, and the smooth trend function $f_C(x) = 1$;
DGP4 Three jumps: $m_0 = 3$ with $d_1 = 2$ and $\lambda_1 = 0.25, d_2 = -2$ and $\lambda_2 = 0.50$, $d_3 = 2$ and $\lambda_3 = 0.75$, and the smooth trend function $f_C(x) = 2\left[\sin(4\pi x) + \sin(8\pi x)\right]$.

We set $n = 512$ and we consider the process of ε_i with Hurst parameter $H = 0.5$ or $H = 0.9$ and $Var(\varepsilon_i) = 0.1$. In Figures 1 and 2, we show the plots of Y_i based on four different trend function and two different Hurst parameters.

In Tables 1-4, we report the estimated number of jumps by our proposed $\widehat{m}$ along with Wang's (1995, 1999) $\widehat{m}^*$. Tables 1 and 2 respectively show the results for DGP1 and DGP2 under $m_0 = 0$. Tables 3 and 4 respectively show the results for DGP3 and 4 under $m_0 = 3$.[9]

There are three practical issues to be considered in the computations of $\widehat{m}$ and $\widehat{m}^*$. The first issue concerns the choice of the wavelet function $\psi(t)$ and vanishing moments M. We find that our simulation results are quite robust to different wavelet functions (Daubechies and Coiflets classes) as long as Assumption 1 holds. Thus, we report only results using the Daubechies wavelet with $M = 4$. With this choice of wavelet function, support is given by $T = 7$, so that Assumption 2 requires the distance between any two break fractions to be greater than $7/2^j$. In DGP3 and DGP4, since the distance between the two break fractions is 0.25, Assumption 2 is satisfied when $j = 6, 7, 8$ are used.

---

[8] In respect to the convergence rate of the jump location estimator with the known m_0, our method reduces to Wang's (1995, 1999) because there is no need to conduct a test. Thus, $\widehat{\lambda}_l$ has the convergence rate of $n/(\log n)^\eta$ where $\eta > 1$ (rate does not depend on H) when $2^j \asymp n/(\log n)^\eta$. Raimondo (1998, Proposition 3.1) provides a modified procedure with the optimal convergence rate n (without logarithm term) for iid data.

[9] In Tables 1 and 3 (the last two rows), we also report the results from the Wild Binary Segmentation (WBS) and the Tail-Greedy Unbalanced Haar (TGUH) of Fryzlewicz (2014, 2018), based on the R package breakfast. In general, both WBS and TGUH methods are designed for a piecewise-constant trend function with short memory noise ($H = 0.5$), so that they offer only reasonable peformances on $f_C(x) = 1$ and $H = 1/2$; otherwise, their performances under other model misspecifications can lead to the severe overestimation of m_0.

The second issue concerns the choice of scale j. In our experiment, we consider different cases regarding the choice of scale j. Because $n = 2^9\,(= 512)$, we can interpret the choice of $j = 6, 7, 8$ corresponding to $2^j = cn^\beta$ with $\beta = 5/9, 6/9, 7/9$ in case of $c = 2$, or with $\beta = 4/9, 5/9, 6/9$ in case of $c = 2^2$. Recall that the range of 2^j specified in Assumption 3* nests $2^j \asymp n^\beta$ with $\beta \in (1/2, 2/3)$ in Assumption 3, so that at least one j's is within the suitable range for computing $\widehat{m}$ and $\widehat{m}^*$. Let us denote $\widehat{m}$ based on the scale j by $\widehat{m}_j$. In addition to reporting the results of $\widehat{m}_j$ for each $j = 6, 7, 8$, we also evaluate the performance of the ensemble estimators given by $\max\{\widehat{m}_7, \widehat{m}_8\}$, $\max\{\widehat{m}_6, \widehat{m}_7\}$, and $\max\{\widehat{m}_6, \widehat{m}_7, \widehat{m}_8\}$. Likewise, in addition to $\widehat{m}_j^*$ for each $j = 6, 7, 8$, we also compute $\max\{\widehat{m}_7^*, \widehat{m}_8^*\}$, $\max\{\widehat{m}_6^*, \widehat{m}_7^*\}$, and $\max\{\widehat{m}_6^*, \widehat{m}_7^*, \widehat{m}_8^*\}$. The choice of single and ensemble estimators is indicated in the first column of Tables 1 to 4.

The third issue concerns on the choice of the nominal size γ used in the sequential testing procedure. The nominal size γ should approach zero as the sample size n increases to establish the consistency of the estimated number of jumps in Theorem 3.2 and Proposition 4. In this experiment with the finite sample, we set γ at 0.05 and 0.01 to compute critical values for $\widehat{m}$ and $\widehat{m}^*$. Note that the critical values for $\widehat{m}$ are $C_{0.05}$ and $C_{0.01}$ from the standard normal distribution, while the critical values $C_{0.05}^*$ and $C_{0.01}^*$ for $\widehat{m}^*$ are obtained from the extreme-value distribution given in Equation (6). In addition, following the recommendation by Wang (1995, 1999), we also employ the critical value $\lim_{\gamma\to 0} C_\gamma^* = \left(2\log 2^j\right)^{1/2} \widetilde{\sigma}_{j,\mathbf{K}}$ for $\widehat{m}^*$. The second columns of Tables 1 to 4 reflect the choice of γ for $\widehat{m}$ and $\widehat{m}^*$. Finally, to enhance the numerical performance, we also compute a translation-invariant vaguelette transformation by cycle-spinning DWT of Y_i (Coifman and Donoho, 1995, Nason and Silverman, 1995). For various combinations of j and γ, we generate $10,000$ realizations of $\widehat{m}$ and $\widehat{m}^*$.

For $m_0 = 0$, the relative frequencies of $\widehat{m}$ (or $\widehat{m}^*$) $= m_0$ and $\widehat{m}$ (or $\widehat{m}^*$) $> m_0$ in $10,000$ replications are reported in Tables 1 and 2 for DGP1 and DGP2, respectively. For each table, Columns 3 and 4 show the results for $H = 0.5$, and Columns 6 and 7 show the results for $H = 0.9$. In addition to the relative frequencies, Column 5 of Tables 1 and 2 show the averages of estimation errors for $H = 0.5$, which corresponds to the bias (namely $E\widehat{m} - m_0$ and $E\widehat{m}^* - m_0$). Similarly, Column 8 of Tables 1 and 2 show the bias for $H = 0.9$.

For $m_0 = 3$, the relative frequencies of $\widehat{m}$ (or $\widehat{m}^*$) $< m_0$, $\widehat{m}$ (or $\widehat{m}^*$) $= m_0$ and $\widehat{m}$ (or $\widehat{m}^*$) $> m_0$ in $10,000$ replications are reported in Tables 3 and 4 for DGP3 and DGP4, respectively. For each table, Columns 3 to 5 show the results for $H = 0.5$, and Columns 7 to 9 show the results for $H = 0.9$. In addition, Column 6 shows the bias under $H = 0.5$, while Column 10 shows the bias for $H = 0.9$.

Overall, the results are very encouraging regarding the performance of our proposed estimator $\widehat{m}$ in comparison with that of Wang's estimator $\widehat{m}^*$. First, the relative frequency of $\widehat{m} = m_0$ is close to 0.95 when $\gamma = 0.05$, and 0.99 when $\gamma = 0.01$ for almost all cases from Table 1 to Table 4. For DGP3 and DGP4, when $m_0 = 3$, the relative frequencies of $\widehat{m} = m_0$ are approaching $1 - \gamma$, and the relative frequencies of $\widehat{m} > m_0$ are approaching γ. In contrast, the relative frequency of $\widehat{m}^* = m_0$ is much lower than $1 - \gamma$ from Table 1 to Table 4. As explained in the previous section, this result is caused by the slow convergence rate of the test statistic $\sup_{k\in\mathbf{K}} \left|\mathbb{W}_{j,k}^{\mathbf{Y}}\right|$, which results in the Type I error being larger than the nominal level γ in the finite sample when the critical values $C_{0.05}^*$, $C_{0.01}^*$ and $\lim_{\gamma\to 0} C_\gamma^*$ are used. Notice that the performances of $\widehat{m}$ and $\widehat{m}^*$ are not very sensitive to the specifications of the smooth trend function $f_C(x)$ $(= 1$ or $2[\sin(4\pi x) + \sin(8\pi x)])$ and changes in the Hurst parameter H $(= 0.5$ or $0.9)$. Second, when the bias of $\widehat{m}$ and $\widehat{m}^*$ are evaluated by the average of estimation errors in $10,000$ replications, the bias of $\widehat{m}$ tends to be much smaller than that of $\widehat{m}^*$ in all cases. For DGP1 and DGP2 with $m_0 = 0$, the medians of the bias of $\widehat{m}$ shown in Tables 1 and 2 are around 0.100, while those of $\widehat{m}^*$ are around 0.701. Similarly, for DGP3 and DGP4 with $m_0 = 3$, the medians of the bias of $\widehat{m}$ shown in Tables 3 and 4 are around 0.151, while those of $\widehat{m}^*$ are around 0.751. Third, from Table 1 to Table 4 we find that choosing a smaller nominal size γ is preferred for $\widehat{m}$ in terms of both the relative frequency of $\widehat{m} = m_0$ and the bias. Fourth, in terms of the relative frequency of $\widehat{m} = m_0$, we find that the performance of single scale estimators is better than the performance of the ensemble estimators of $\widehat{m}$. When $m_0 = 0$, Tables 1 and 2 show that using the single estimators of $\widehat{m}$ makes the relative frequency of $\widehat{m} = m_0$ closer to $1 - \gamma$ than using the ensemble estimators. Moreover, when $m_0 = 3$, Tables 3 and 4 show that the single scale estimators of $\widehat{m}$ are also preferred because the ensemble estimators tend to have a higher relative frequency of over-estimating the number of jumps ($\widehat{m} > m_0$).[10] In contrast, regardless of jump sizes, Wang's ensemble estimators $\widehat{m}^*$ always have higher relative frequencies of overestimating the number of jumps ($\widehat{m}^* > m_0$) than Wang's single scale estimators.

6. Applications. In what follows, we estimate the number of jumps and jump locations in two groups of time series data, the stock price index and Nile river data. For our proposed estimator of the number of jumps $\widehat{m}$ and jump locations $\left\{\widehat{\lambda}_l\right\}_{l=1}^{\widehat{m}}$, the critical values from either $C_{0.05}$ or $C_{0.01}$ based on the quantile function of the standard Gaussian variable are used. For Wang's (1995, 1999) estimator of the number of jumps $\widehat{m}^*$ and locations $\left\{\widehat{\lambda}_l^*\right\}_{l=1}^{\widehat{m}^*}$, $C^*_{0.05}$, $C^*_{0.01}$ and $\lim_{\gamma\to 0} C^*_\gamma = \left(2\log 2^j\right)^{1/2} \widetilde{\sigma}_{j,\mathbf{K}}$ are used as critical values. The scale j is to chosen to be 7 for both the stock price and the Nile River cases where the sample size $n = 512$. Recall that with the choice of $j = 7$ and the Daubechies wavelet function with $M = 4$, we are assuming that distances between break fractions are at least greater than $7/128$ $(= T/2^j)$. The translation-invariant vaguelette transformation is also applied to compute $\mathbb{W}^{\mathbf{Y}}_{j,k}$. When jumps are detected, we further estimate the nonlinear trend functions for each interval between the estimated jump locations by the local quadratic regression estimator available in the locpoly function of the R package KernSmooth. Moreover, for each interval segmented by $\left\{\widehat{\lambda}_l\right\}_{l=1}^{\widehat{m}_7}$, we estimate the Hurst parameter H by the wavelet regression estimator available in the WVLM function of the R package WaveLetLongMemory.

6.1. *Daily Dow-Jones Industrial Average.* For the first application, we estimate the jump in the US stock price index using the logarithm daily series of the Dow-Jones Industrial Average at market close from 2019-08-22 to 2021-09-01 where the sample size n is 512. The data is obtained from the FRED database compiled by the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/series/DJIA).

In Table 5 we demonstrate how sequential procedures are employed to estimate the number of jumps. In what follows, we use $C_{0.01}$ for the critical value of $\widehat{D}_{7,\mathbf{K}}$ and $\lim_{\gamma\to 0} C^*_\gamma$ for the critical value of $\sup_{k\in\mathbf{K}} \left|\mathbb{W}^{\mathbf{Y}}_{7,k}\right|$ because the simulations provided in the previous section suggested that the use of $C_{0.01}$ and $\lim_{\gamma\to 0} C^*_\gamma$ provided better performances in selecting of the true number of jumps, in comparison with other choices of critical values. For the sequential procedure based on $\widehat{D}_{7,\mathbf{K}}$: we reject the first two null hypotheses of $H_0 : m_0 = 0$ and $H_0 : m_0 = 1$ because the corresponding values $\widehat{D}_{7,\mathbf{K}} = 4.1204$ and 3.8619 are greater than the critical value $C_{0.01} = 2.5788$ (or $C_{0.05} = 1.9600$). However, we fail to reject the null hypothesis $H_0 : m_0 = 2$ because the corresponding test statistic $\widehat{D}_{7,\mathbf{K}} = 2.2824$ is less

[10]However, exceptions are that ensemble estimators $\max\{\widehat{m}_7, \widehat{m}_8\}$ and $\max\{\widehat{m}_6, \widehat{m}_7, \widehat{m}_8\}$ under DGP3 and DGP4 are better than the single scale estimator $\widehat{m}_8$.

than $C_{0.01} = 2.5788$, and thus $\widehat{m}_7 = 2$ is obtained. In contrast, for Wang's (1995,1999) sequential procedure, we reject the null hypothesis $H_0 : m_0 = 2$ but fail to reject the null hypothesis $H_0 : m_0 = 3$. From the table, the test statistic $\sup_{k\in\mathbf{K}} \left|\mathbb{W}^{\mathbf{Y}}_{7,k}\right| = 0.002688$ for $H_0 : m_0 = 2$ is greater than the critical value $\lim_{\gamma\to 0} C^*_\gamma = 0.001308$, while the test statistic $\sup_{k\in\mathbf{K}} \left|\mathbb{W}^{\mathbf{Y}}_{7,k}\right| = 0.000940$ for $H_0 : m_0 = 3$ is less than the critical value $\lim_{\gamma\to 0} C^*_\gamma = 0.001308$. Therefore, $\widehat{m}^*_7 = 3$ is obtained. The fact that the sequential procedure based on $\widehat{D}_{7,\mathbf{K}}$ detected fewer jumps compared to that of Wang's (1995,1999) sequential procedure, which is based on $\sup_{k\in\mathbf{K}} \left|\mathbb{W}^{\mathbf{Y}}_{7,k}\right|$, is consistent with the discussion in Sections 4 and the simulation study of Section 5. From Table 5 we also observe that if we increase the nominal size γ (the corresponding critical values become smaller), both $\widehat{m}_7$ and $\widehat{m}^*_7$ become larger. This result occurs because, when the corresponding critical values become smaller, rejecting the null hypothesis becomes easier.

Panel (a) of Figure 3 shows that estimated jump locations based on our procedure are 2020-03-20 and 2020-06-10 ($\widehat{m}_7 = 2$). Panel (b) of Figure 3 shows that estimated jump locations based on Wang's procedure are 2020-03-20, 2020-05-12 and 2020-06-10 ($\widehat{m}^*_7 = 3$). Notice that 2020-03-20 corresponded to the date of the World Health Organization issuing the Covid-19 health alert, and 2020-06-10 was the date when the number of confirmed cases of COVID-19 in the United States hit two million. We also estimate the Hurst parameter H for each segment in which the segment is based on $\left\{\widehat{\lambda}_l\right\}_{l=1}^{\widehat{m}_7}$ where $\widehat{m}_7 = 2$. Our estimates of the Hurst parameter H from the first to the third segment are 0.56**,** 0.63 and 0.77, respectively. Since all the estimates are greater than 0.5, our findings have highlighted the US stock market has a long memory feature even though we took into account of jumps in the trend function.

6.2. *The Nile River.* For the second application, we estimate jumps in the minimum yearly water level of the Nile River. The data is available from years 662 to 1284 in Page 237 of Beran (1994), and we use the observations from years 773 to 1284 so that the sample size n is 512.[11]

Table 6 presents the sequential procedures in estimating the number of jumps that are similar to the application of the Dow-Jones Industrial Average data. For the sequential procedure based on $\widehat{D}_{7,\mathbf{K}}$, we reject the first three null hypotheses of $H_0 : m_0 = 0$, $H_0 : m_0 = 1$ and $H_0 : m_0 = 2$ because the corresponding values $\widehat{D}_{7,\mathbf{K}} = 4.4520$, 3.6423 and 2.9267 are greater than the critical value $C_{0.01} = 2.5788$ (or $C_{0.05} = 1.9600$). However, we fail to reject the null hypothesis $H_0 : m_0 = 3$ because the corresponding test statistic $\widehat{D}_{7,\mathbf{K}} = 2.3435$ is less than $C_{0.01} = 2.5788$, and thus $\widehat{m}_7 = 3$ is obtained. In contrast, for Wang's (1995,1999) sequential procedure, we reject the null hypothesis $H_0 : m_0 = 3$ but fail to reject the null hypothesis $H_0 : m_0 = 4$. From the table, the test statistic $\sup_{k\in\mathbf{K}} \left|\mathbb{W}^{\mathbf{Y}}_{7,k}\right| = 9.6761$ for $H_0 : m_0 = 3$ is greater than the critical value $\lim_{\gamma\to 0} C^*_\gamma = 7.2975$, while the test statistic $\sup_{k\in\mathbf{K}} \left|\mathbb{W}^{\mathbf{Y}}_{7,k}\right| = 7.1524$ for $H_0 : m_0 = 4$ is less than the critical value $\lim_{\gamma\to 0} C^*_\gamma = 7.2975$. Therefore, $\widehat{m}^*_7 = 4$ is obtained. Notice that again the selected number of jumps based on our procedure tends to be less than that based on Wang's, which can be explained by the same reason as that used in the stock price application.

Panel (a) of Figure 4 shows that the estimated jump locations based on our procedure are years 780, 962 and 1032 ($\widehat{m}_7 = 3$). Panel (b) of Figure 4 shows that estimated jump locations

[11]We can easily adapt our procedure to work with time series that are not multiples of the power of two by padding as suggested in Chapter 4.11 of Percival and Walden (2000), as is true here ($n = 663$).

based on Wang's procedure are years 780, 884, 962 and 1032 ($\widehat{m}_7^* = 4$). We estimate the Hurst parameter H for each segment of Panel (a) in which the segment is based on $\left\{\widehat{\lambda}_l\right\}_{l=1}^{\widehat{m}_7}$ where $\widehat{m}_7 = 3$. The estimated Hurst parameters H for the second, third and fourth segments are 0.69, 0.77 and 0.55, respectively, which indicate the long memory feature of the Nile River data. In contrast, these Hurst parameter estimates are less than the estimated value 0.95 obtained by Percival and Walden (2000). This difference may come from the fact that Percival and Walden (2000) did not consider the possibility of potential jumps in the trend function.[12]

7. Conclusion. In this paper, we propose a new inference procedure to estimate the number of jumps and their locations. One notable feature of our proposed procedure is that the construction of the test statistic is based on the wavelet variance instead of the supremum of wavelet coefficients as in Wang (1995, 1999). As a result, our test statistic, converges to the normal distribution, compared to Wang's sup-type test statistic converging very slowly to an extreme-value distribution. In simulation experiments, we demonstrate that our procedure works better than Wang's procedure in a finite sample.

Although the current paper focuses on the noise ε_i belonging to the so-called fractionally integrated white noise process, ARFIMA$(0, H, 0)$ (independent or fractional Gaussian error), it is possible to extend ε_i to the general ARFIMA(p, H, q) process. In particular, our procedure would still apply under this general setup. However, it would be difficult to apply Wang's procedure because the analytical critical value of the extreme-value distribution under the general dependence is not covered under Wang (1995, 1999), and its critical value depends on the knowledge of p and q, which need to be estimated separately (Leadbetter, Lindgren and Rootzén, 1983).[13] In contrast, our test statistic converges to the standard normal distribution without the needing of estimating p and q.

Acknowledgments. We thank participants at the 2021 NBER-NSF Time Series Conference for helpful comments. We thank Yanqin Fan, Rafal Kuilk, Matthew Strathearn, Tatsushi Oka and Yohei Yamamoto for helpful comments and discussions. We thank Kim Huynh for his continuous support and encouragement. The use of the Bank of Canada EDITH High Performance cluster is gratefully acknowledged. All errors are our own.

[12] Note that our focus is on testing the existence of jumps, rather than testing the long-memory versus a spurious one. For the latter, see Diebold and Inoue (2001), Shimotsu (2006), Perron and Qu (2010), Iacone (2010), Baek and Pipiras (2012), McCloskey and Perron (2013) and Hou and Perron (2014). The above mentioned papers differentiate between a long-range dependent time series and a weakly dependent time series with a change-point in the mean. However, our target is to detect jumps in which a genuine long-memory error could exist for both the null and alternative hypotheses.

[13] By taking advantage of the strong invariance principle for the short-memory noise with $H = 0.5$, Wu and Zhao (2007) provide a simulation method to obtain the critical value of the extreme-value distribution under ARFIMA(p, H, q) where $H = 0.5$ or ARMA(p, q). However, Wu and Zhao's (2007) approach is not applicable to the long memory case when $H > 0.5$.

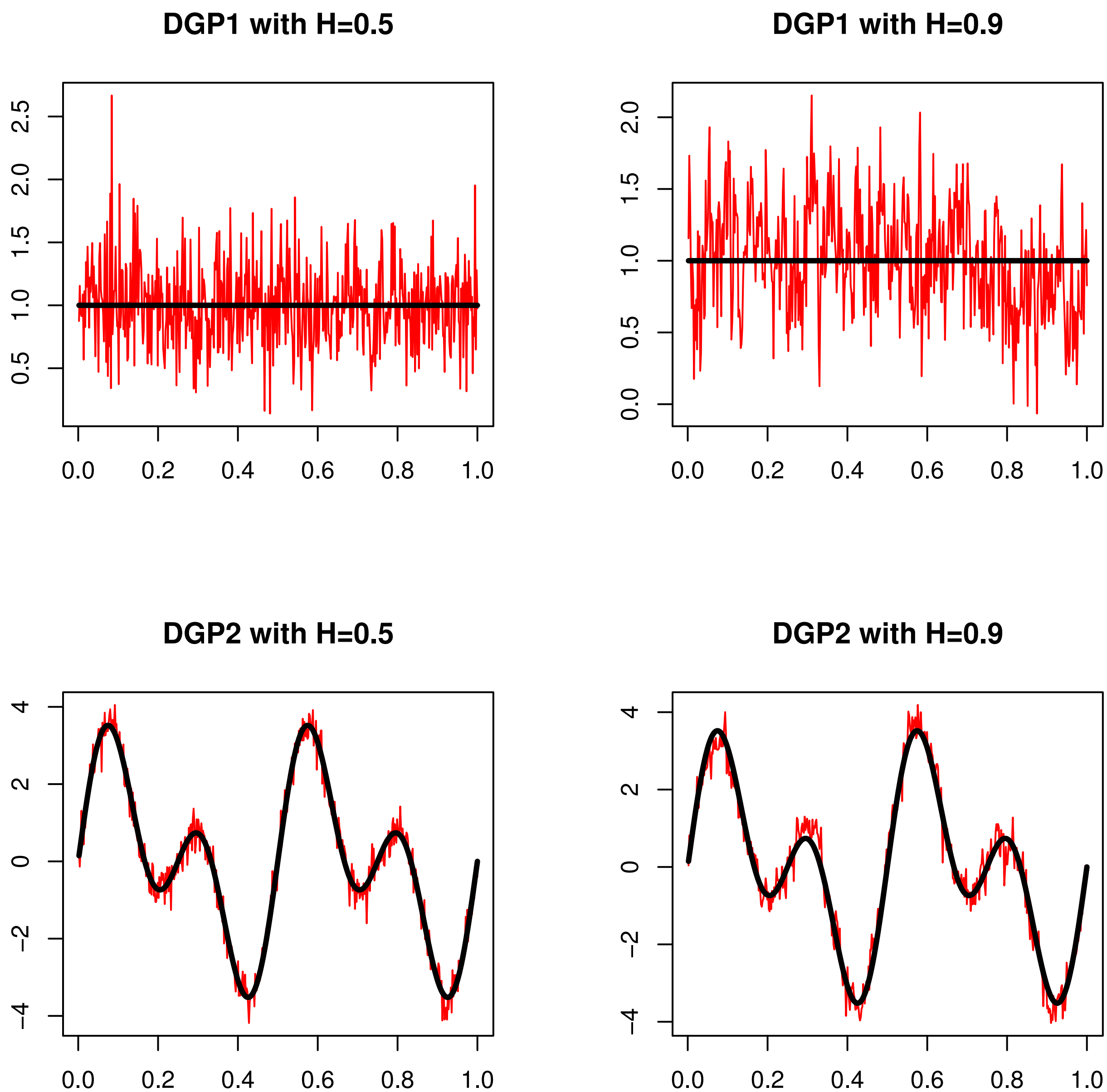

FIG 1. *DGP1 refers to* $m_0 = 0$ *and the smooth trend function* $f_C(x) = 1$. *DGP2 refers to* $m_0 = 0$ *and the smooth trend function* $f_C(x) = 2\left[\sin(4\pi x) + \sin(8\pi x)\right]$. H *is the Hurst parameter.*

REFERENCES

[1] Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica: Journal of the Econometric Society, 817-858.

[2] Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica: Journal of the Econometric Society, 1383-1414.

[3] Ayache, A., & Bertrand, P. R. (2011). Discretization error of wavelet coefficient for fractal like processes. Advances in Pure and Applied Mathematics, 2(2), 297-321.

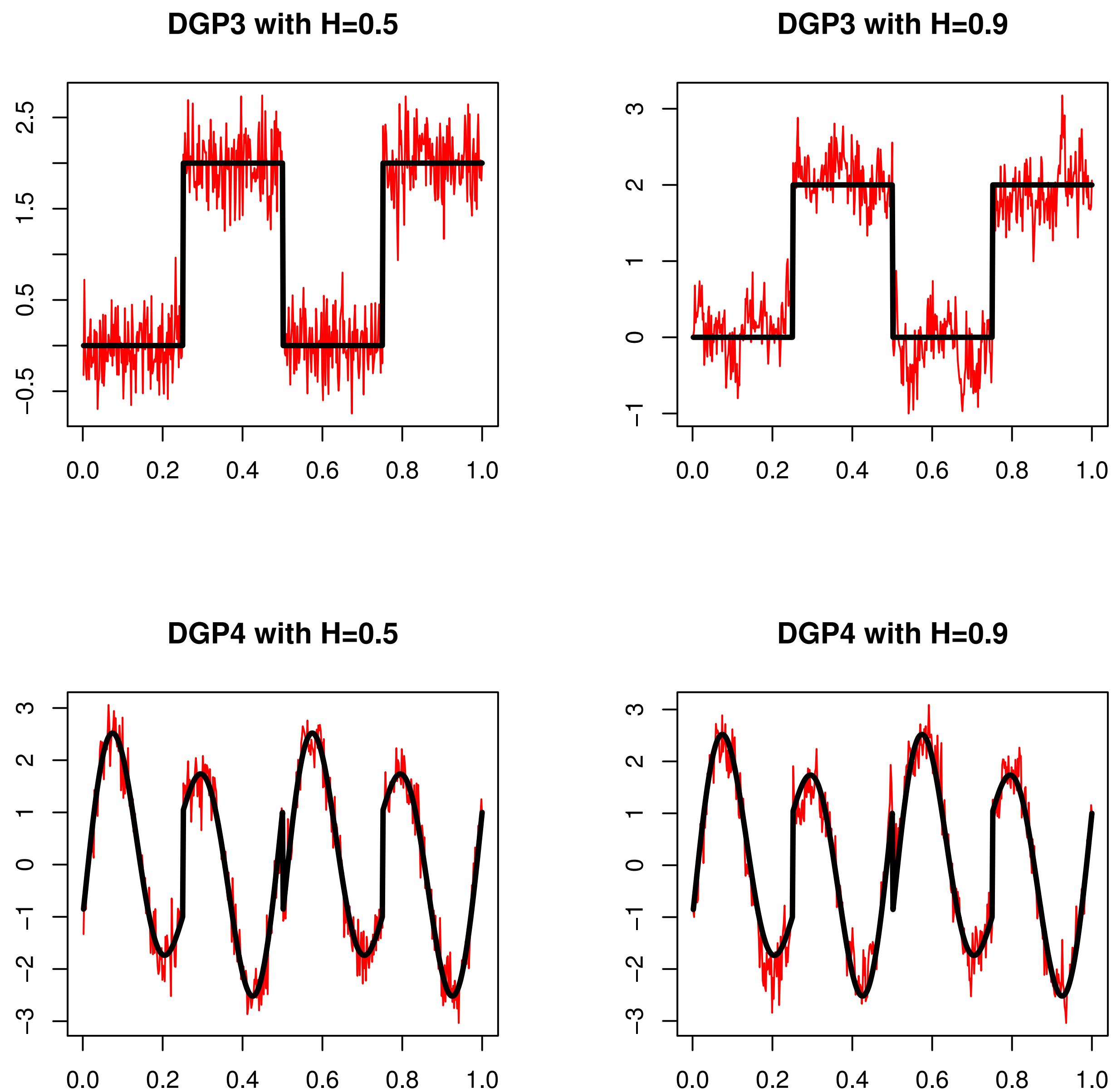

FIG 2. *DGP3 refers to* $m_0 = 3$ *with* $d_1 = 2$ *and* $\lambda_1 = 0.25, d_2 = -2$ *and* $\lambda_2 = 0.50$, $d_3 = 2$ *and* $\lambda_3 = 0.75$, *and the smooth trend function* $f_C(x) = 1$. *DGP4 refers to* $m_0 = 3$ *with* $d_1 = 2$ *and* $\lambda_1 = 0.25, d_2 = -2$ *and* $\lambda_2 = 0.50$, $d_3 = 2$ *and* $\lambda_3 = 0.75$, *and the smooth trend function* $f_C(x) = 2\left[\sin(4\pi x) + \sin(8\pi x)\right]$. H *is the Hurst parameter.*

[4] Bai, J. (1997), Estimating multiple breaks one at a time, Econometric Theory, 13, 315-352.

[5] Bai, J., and Perron, P. (1998), Estimating and testing linear models with multiple structural changes, Econometrica, 66, 47-78.

[6] Baek, C., & Pipiras, V. (2012). Statistical tests for a single change in mean against long-range dependence. Journal of Time Series Analysis, 33(1), 131-151.

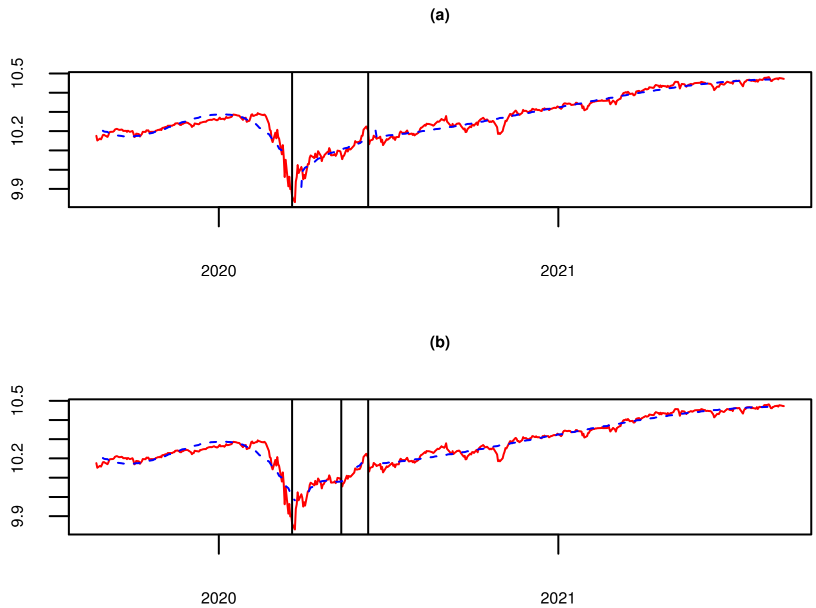

FIG 3. *The daily series of the Dow-Jones Industrial Average at market close in logarithm from 2019-08-22 to 2021-09-01 (solid line). Panel (a) shows two jump locations* $\{2020\text{-}03\text{-}20, 2020\text{-}06\text{-}10\}$ *(vertical lines) obtained from* $\left\{\widehat{\lambda}_l\right\}_{l=1}^{\widehat{m}}$ *using the critical value* $C_{0.01}$*. Panel (b) shows three jump locations* $\{2020\text{-}03\text{-}20, 2020\text{-}05\text{-}12, 2020\text{-}06\text{-}10\}$ *(vertical lines) obtained from* $\left\{\widehat{\lambda}_l^*\right\}_{l=1}^{\widehat{m}^*}$ *using the critical value* $\lim_{\gamma\to 0} C_\gamma^*$*. For both panels, the fitted trends using a piecewise local quadratic regression (dashed line) are also shown.*

[7] Beran, J. (1994). Statistics for Long-Memory Processes (1st ed.). Routledge. https://doi.org/10.1201/9780203738481.

[8] Beran, J., & Shumeyko, Y. (2012). Bootstrap testing for discontinuities under long-range dependence. Journal of Multivariate Analysis, 105(1), 322-347.

[9] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. The Annals of Statistics, 14(3), 1171-1179.

[10] Chen, G., Y. K. Choi, and Y. Zhou (2008). Detections of changes in return by a wavelet smoother with conditional heteroskedastic volatility. Journal of Econometrics 143, 227-262.

[11] Chen, H., & Fan, Y. (2019). Identification and wavelet estimation of weighted ATE under discontinuous and kink incentive assignment mechanisms. Journal of Econometrics, 212(2), 476-502.

[12] Coifman, R. R., & Donoho, D. L. (1995). Translation-invariant de-noising. In Wavelets and Statistics, 125-150. Springer, New York, NY.

[13] Cranston, M., Scheutzow, M., & Steinsaltz, D. (2000). Linear bounds for stochastic dispersion. The Annals of Probability, 28(4), 1852-1869.

[14] Daubechies, I. (1992). Ten lectures on wavelets. Society for Industrial and Applied Mathematics.

[15] Diebold, F. X., & Inoue, A. (2001). Long memory and regime switching. Journal of econometrics, 105(1), 131-159.

[16] Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81(3), 425-455.

[17] Donoho, D. L., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association, 90(432), 1200-1224.

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

[19] Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. The Annals of Statistics, 42(6), 2243-2281.

[20] Fryzlewicz, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. The Annals of Statistics, 46(6B), 3390-3421.

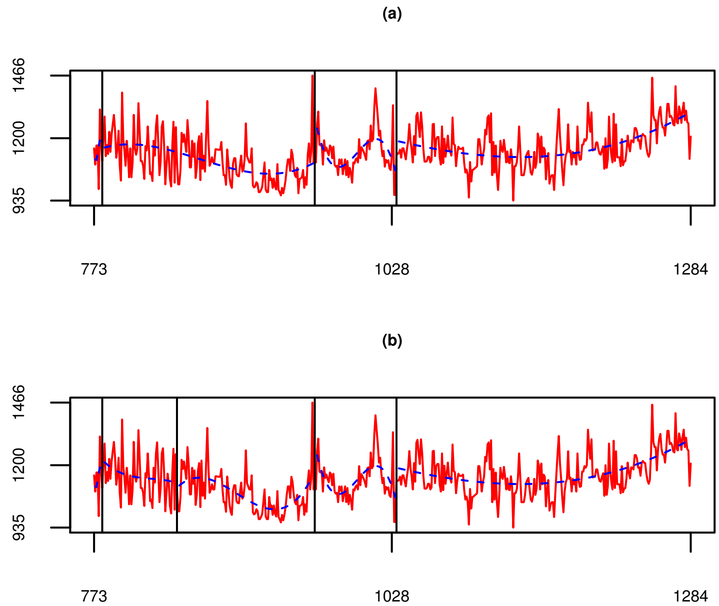

FIG 4. *The annual series of the minimum water level of the Nile River from 773 to 1284 (solid line). Panel (a) shows three jump locations* $\{780, 962, 1032\}$ *(vertical lines) obtained from* $\left\{\widehat{\lambda}_l\right\}_{l=1}^{\widehat{m}}$ *using the critical value* $C_{0.01}$. *Panel (b) shows four jump locations* $\{780, 844, 962, 1032\}$ *(vertical lines) obtained from* $\left\{\widehat{\lambda}_l^*\right\}_{l=1}^{\widehat{m}^*}$ *using the critical value* $\lim_{\gamma\to 0} C_\gamma^*$. *For both panels, the fitted trends using a piecewise local quadratic regression (dashed line) are also shown.*

[21] Gao, J., Gijbels, I., & Van Bellegem, S. (2008). Nonparametric simultaneous testing for structural breaks. Journal of Econometrics, 143(1), 123-142.

[22] Gijbels, I., & Goderniaux, A. C. (2004). Bandwidth selection for change point estimation in nonparametric regression. Technometrics, 46(1), 76-86.

[23] Hall, P. (1979). On the rate of convergence of normal extremes. Journal of Applied Probability, 16(2), 433-439.

[24] Hall, P., & Hart, J. D. (1990). Nonparametric regression with long-range dependence. Stochastic Processes and their Applications, 36(2), 339-351.

[25] Härdle, W., Kerkyacharian, G., Picard, D., & Tsybakov, A. (2012). Wavelets, approximation, and statistical applications, 129. Springer Science & Business Media.

[26] Hou, J., & Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. Journal of Econometrics, 182(2), 309-328.

[27] Iacone, F. (2010). Local Whittle estimation of the memory parameter in presence of deterministic components. Journal of Time Series Analysis, 31(1), 37-49.

[28] Kouamo, O., Lėy-Leduc, C., & Moulines, E. (2013). Central limit theorem for the robust log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context. Bernoulli, 19(1), 172-204.

[29] Krämer, W., Sibbertsen, P., & Kleiber, C. (2002). Long memory vs. structural change in financial time series. Allgemeines Statistisches Archive, 86, 83-96.

[30] Lavielle, M., & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. Journal of Time Series Analysis, 21(1), 33-59.

[31] Lazarová, Š. (2005). Testing for structural change in regression with long memory processes. Journal of Econometrics, 129(1-2), 329-372.

[32] Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). Extremes and related properties of random sequences and processes. Springer Series in Statistics.

[33] Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M. S., & Reisen, V. A. (2011). Robust estimation of the scale and of the autocovariance function of Gaussian short-and long-range dependent processes. Journal of Time Series Analysis, 32(2), 135-156.

TABLE 1
Relative frequency of estimates of the number of jumps under the smooth trend function $f_C(x) = 1$ and no jump $m_0 = 0$

		$H = 0.5$			$H = 0.9$		
		Relative frequency		Bias	Relative frequency		Bias
		$= m_0$	$> m_0$		$= m_0$	$> m_0$	
(1) Sequential procedure based on $\widehat{D}_{j,\mathbf{K}}$							
$\widehat{m}_8$	$\gamma = 0.05$	0.958	0.042	0.178	0.961	0.039	0.164
	$\gamma = 0.01$	0.987	0.013	0.048	0.987	0.013	0.044
$\widehat{m}_7$	$\gamma = 0.05$	0.953	0.047	0.162	0.952	0.048	0.166
	$\gamma = 0.01$	0.983	0.017	0.052	0.985	0.015	0.046
$\widehat{m}_6$	$\gamma = 0.05$	0.948	0.052	0.140	0.947	0.053	0.168
	$\gamma = 0.01$	0.982	0.018	0.046	0.979	0.021	0.055
$\max\{\widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.913	0.087	0.336	0.914	0.086	0.328
	$\gamma = 0.01$	0.970	0.030	0.100	0.973	0.027	0.090
$\max\{\widehat{m}_6, \widehat{m}_7\}$	$\gamma = 0.05$	0.904	0.096	0.297	0.901	0.099	0.330
	$\gamma = 0.01$	0.965	0.035	0.098	0.965	0.035	0.101
$\max\{\widehat{m}_6, \widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.865	0.135	0.470	0.865	0.135	0.486
	$\gamma = 0.01$	0.952	0.048	0.146	0.952	0.048	0.144
(2) Wang's (1995,1999) sequential procedure based on $\sup_{k\in\mathbf{K}} \left\| \mathbb{W}^{\mathbf{Y}}_{j,k} \right\|$							
$\widehat{m}^*_8$	$\gamma = 0.05$	0.715	0.285	0.347	0.752	0.248	0.295
	$\gamma = 0.01$	0.728	0.272	0.328	0.763	0.237	0.278
	$\gamma \to 0$	0.782	0.218	0.253	0.820	0.180	0.204
$\widehat{m}^*_7$	$\gamma = 0.05$	0.627	0.373	0.508	0.600	0.400	0.556
	$\gamma = 0.01$	0.642	0.358	0.479	0.618	0.382	0.522
	$\gamma \to 0$	0.854	0.146	0.164	0.859	0.142	0.158
$\widehat{m}^*_6$	$\gamma = 0.05$	0.343	0.657	1.269	0.304	0.696	1.427
	$\gamma = 0.01$	0.363	0.637	1.205	0.320	0.680	1.361
	$\gamma \to 0$	0.831	0.169	0.203	0.828	0.171	0.205
$\max\{\widehat{m}^*_7, \widehat{m}^*_8\}$	$\gamma = 0.05$	0.452	0.548	0.740	0.454	0.546	0.743
	$\gamma = 0.01$	0.471	0.529	0.701	0.474	0.526	0.703
	$\gamma \to 0$	0.669	0.331	0.383	0.704	0.296	0.336
$\max\{\widehat{m}^*_6, \widehat{m}^*_7\}$	$\gamma = 0.05$	0.218	0.782	1.492	0.183	0.817	1.653
	$\gamma = 0.01$	0.235	0.765	1.423	0.198	0.802	1.579
	$\gamma \to 0$	0.711	0.289	0.341	0.713	0.287	0.337
$\max\{\widehat{m}^*_6, \widehat{m}^*_7, \widehat{m}^*_8\}$	$\gamma = 0.05$	0.158	0.842	1.591	0.138	0.862	1.726
	$\gamma = 0.01$	0.173	0.827	1.522	0.151	0.849	1.652
	$\gamma \to 0$	0.558	0.442	0.527	0.584	0.416	0.489
(3) Fryzlewicz's (2014) based on Wild Binary Segmentation (WBS) and Fryzlewicz's (2018) based on Tail-Greedy Unbalanced Haar (TGUH)							
WBS		0.876	0.124	0.290	–	–	–
TGUH		0.875	0.125	0.290	–	–	–

[34] Ma, Y., & Genton, M. G. (2000). Highly robust estimation of the autocovariance function. Journal of Time Series Analysis, 21(6), 663-684.

[35] McCloskey, A., & Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. Econometric Theory, 29(06), 1196-1237.

TABLE 2
Relative frequency of estimates of the number of jumps under the smooth trend function $f_C(x) \equiv 2[\sin(4\pi x) + \sin(8\pi x)]$ *and no jump* $m_0 = 0$

		$H = 0.5$			$H = 0.9$		
		Relative frequency		Bias	Relative frequency		Bias
		$= m_0$	$> m_0$		$= m_0$	$> m_0$	
(1) Sequential procedure based on $\widehat{D}_{j,\mathbf{K}}$							
$\widehat{m}_8$	$\gamma = 0.05$	0.957	0.043	0.157	0.959	0.041	0.139
	$\gamma = 0.01$	0.984	0.016	0.044	0.986	0.014	0.041
$\widehat{m}_7$	$\gamma = 0.05$	0.946	0.064	0.157	0.950	0.050	0.161
	$\gamma = 0.01$	0.981	0.019	0.047	0.979	0.021	0.056
$\widehat{m}_6$	$\gamma = 0.05$	0.942	0.058	0.148	0.947	0.053	0.125
	$\gamma = 0.01$	0.974	0.026	0.058	0.980	0.020	0.044
$\max\{\widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.905	0.095	0.312	0.911	0.089	0.297
	$\gamma = 0.01$	0.966	0.034	0.091	0.966	0.034	0.097
$\max\{\widehat{m}_6, \widehat{m}_7\}$	$\gamma = 0.05$	0.892	0.108	0.299	0.900	0.100	0.281
	$\gamma = 0.01$	0.956	0.044	0.104	0.960	0.040	0.099
$\max\{\widehat{m}_6, \widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.853	0.147	0.448	0.864	0.136	0.412
	$\gamma = 0.01$	0.941	0.059	0.147	0.947	0.053	0.139
(2) Wang's (1995,1999) sequential procedure based on $\sup_{k\in\mathbf{K}} \left\lvert \mathbb{W}^{\mathbf{Y}}_{j,k} \right\rvert$							
$\widehat{m}^*_8$	$\gamma = 0.05$	0.708	0.292	0.364	0.737	0.263	0.318
	$\gamma = 0.01$	0.720	0.280	0.345	0.750	0.250	0.299
	$\gamma \to 0$	0.775	0.225	0.268	0.811	0.189	0.217
$\widehat{m}^*_7$	$\gamma = 0.05$	0.607	0.393	0.560	0.590	0.410	0.598
	$\gamma = 0.01$	0.624	0.376	0.525	0.606	0.394	0.567
	$\gamma \to 0$	0.840	0.160	0.192	0.836	0.164	0.194
$\widehat{m}^*_6$	$\gamma = 0.05$	0.375	0.625	1.181	0.333	0.667	1.312
	$\gamma = 0.01$	0.392	0.608	1.130	0.348	0.652	1.259
	$\gamma \to 0$	0.825	0.175	0.226	0.826	0.174	0.223
$\max\{\widehat{m}^*_7, \widehat{m}^*_8\}$	$\gamma = 0.05$	0.430	0.570	0.800	0.436	0.564	0.800
	$\gamma = 0.01$	0.451	0.549	0.757	0.457	0.543	0.758
	$\gamma \to 0$	0.652	0.348	0.423	0.678	0.322	0.379
$\max\{\widehat{m}^*_6, \widehat{m}^*_7\}$	$\gamma = 0.05$	0.230	0.770	1.447	0.196	0.804	1.581
	$\gamma = 0.01$	0.248	0.752	1.385	0.211	0.789	1.520
	$\gamma \to 0$	0.693	0.307	0.389	0.692	0.308	0.386
$\max\{\widehat{m}^*_6, \widehat{m}^*_7, \widehat{m}^*_8\}$	$\gamma = 0.05$	0.163	0.837	1.561	0.146	0.854	1.663
	$\gamma = 0.01$	0.179	0.821	1.497	0.160	0.840	1.598
	$\gamma \to 0$	0.537	0.463	0.586	0.560	0.440	0.544

[36] Mémin, J., Mishura, Y., & Valkeila, E. (2001). Inequalities for the moments of Wiener integrals with respect to a fractional Brownian motion. Statistics & Probability Letters, 51(2), 197-206.

[37] Müller, H. G. (1992). Change-points in nonparametric regression analysis. The Annals of Statistics, 737-761.

[38] Müller, H. G., & Stadtmüller, U. (1999). Discontinuous versus smooth regression. The Annals of Statistics, 27(1), 299-337.

[39] Nason, G. P., & Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. In Wavelets and Statistics, 281-299. Springer, New York, NY.

TABLE 3
Relative frequency of estimates of the number of jumps under the smooth trend function $f_C(x) = 1$ and three jumps $m_0 = 3$ with $d_1 = 2$, $d_2 = -2$, and $d_3 = 2$

		$H = 0.5$				$H = 0.9$			
		Relative frequency			Bias	Relative frequency			Bias
		$< m_0$	$= m_0$	$> m_0$		$< m_0$	$= m_0$	$> m_0$	
(1) Sequential procedure based on $\widehat{D}_{j,\mathbf{K}}$									
$\widehat{m}_8$	$\gamma = 0.05$	0.266	0.691	0.043	-0.087	0.036	0.925	0.039	0.124
	$\gamma = 0.01$	0.449	0.539	0.012	-0.412	0.098	0.889	0.013	-0.050
$\widehat{m}_7$	$\gamma = 0.05$	0.000	0.953	0.047	0.163	0.000	0.953	0.047	0.166
	$\gamma = 0.01$	0.001	0.983	0.016	0.052	0.001	0.984	0.015	0.05
$\widehat{m}_6$	$\gamma = 0.05$	0.001	0.947	0.052	0.145	0.001	0.945	0.054	0.161
	$\gamma = 0.01$	0.001	0.978	0.021	0.050	0.001	0.980	0.019	0.052
$\max\{\widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.000	0.913	0.087	0.340	0.000	0.916	0.084	0.322
	$\gamma = 0.01$	0.001	0.971	0.028	0.099	0.000	0.972	0.028	0.098
$\max\{\widehat{m}_6, \widehat{m}_7\}$	$\gamma = 0.05$	0.000	0.903	0.097	0.307	0.000	0.901	0.099	0.323
	$\gamma = 0.01$	0.000	0.964	0.036	0.105	0.000	0.966	0.034	0.103
$\max\{\widehat{m}_6, \widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.000	0.864	0.136	0.482	0.000	0.866	0.134	0.475
	$\gamma = 0.01$	0.000	0.950	0.050	0.151	0.000	0.953	0.047	0.149
(2) Wang's (1995,1999) sequential procedure based on $\sup_{k\in\mathbf{K}} \left\| \mathbb{W}^{\mathbf{Y}}_{j,k} \right\|$									
$\widehat{m}^*_8$	$\gamma = 0.05$	0.000	0.737	0.263	0.314	0.000	0.773	0.227	0.264
	$\gamma = 0.01$	0.000	0.749	0.251	0.297	0.000	0.786	0.214	0.247
	$\gamma \to 0$	0.000	0.799	0.201	0.230	0.000	0.835	0.165	0.184
$\widehat{m}^*_7$	$\gamma = 0.05$	0.000	0.695	0.305	0.389	0.000	0.673	0.327	0.427
	$\gamma = 0.01$	0.000	0.710	0.290	0.366	0.000	0.688	0.312	0.401
	$\gamma \to 0$	0.000	0.887	0.113	0.125	0.000	0.889	0.111	0.121
$\widehat{m}^*_6$	$\gamma = 0.05$	0.000	0.515	0.485	0.761	0.000	0.471	0.529	0.872
	$\gamma = 0.01$	0.000	0.535	0.465	0.719	0.000	0.486	0.514	0.830
	$\gamma \to 0$	0.000	0.900	0.100	0.113	0.000	0.898	0.102	0.115
$\max\{\widehat{m}^*_7, \widehat{m}^*_8\}$	$\gamma = 0.05$	0.000	0.515	0.485	0.617	0.000	0.521	0.479	0.613
	$\gamma = 0.01$	0.000	0.535	0.465	0.585	0.000	0.542	0.458	0.579
	$\gamma \to 0$	0.000	0.711	0.289	0.329	0.000	0.743	0.257	0.286
$\max\{\widehat{m}^*_6, \widehat{m}^*_7\}$	$\gamma = 0.05$	0.000	0.356	0.645	0.991	0.000	0.316	0.684	1.108
	$\gamma = 0.01$	0.000	0.378	0.622	0.940	0.000	0.335	0.665	1.055
	$\gamma \to 0$	0.000	0.798	0.202	0.226	0.000	0.799	0.201	0.224
$\max\{\widehat{m}^*_6, \widehat{m}^*_7, \widehat{m}^*_8\}$	$\gamma = 0.05$	0.000	0.262	0.738	1.124	0.000	0.244	0.756	1.208
	$\gamma = 0.01$	0.000	0.283	0.717	1.072	0.000	0.261	0.739	1.153
	$\gamma \to 0$	0.000	0.642	0.358	0.411	0.000	0.668	0.332	0.374
(3) Fryzlewicz's (2014) based on Wild Binary Segmentation (WBS) and Fryzlewicz's (2018) based on Tail-Greedy Unbalanced Haar (TGUH)									
WBS		0.000	0.819	0.181	0.386	–	–	–	–
TGUH		0.000	0.815	0.185	0.394	–	–	–	–

[40] Papoulis, A. (1984). Probability, Random Variables and Stochastic Processes (2nd ed.), New York: McGraw-Hill.

[41] Percival, D. P. (1995). On estimation of the wavelet variance. Biometrika, 82(3), 619-631.

[42] Percival, D. B., & Walden, A. T. (2000). Wavelet methods for time series analysis (Vol. 4). Cambridge University Press.

[43] Perron, P., & Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. Journal of Business & Economic Statistics, 28(2), 275-290.

[44] Perron, P., Shintani, M., & Yabu, T. (2020). Trigonometric trend regressions of unknown frequencies with stationary or integrated noise (No. WP2020-012). -Department of Economics, Boston University.

TABLE 4
Relative frequency of estimates of the number of jumps under the smooth trend function $f_C(x) \equiv 2[\sin(4\pi x) + \sin(8\pi x)]$ *and three jumps* $m_0 = 3$ *with with* $d_1 = 2$*,* $d_2 = -2$*, and* $d_3 = 2$

		$H = 0.5$				$H = 0.9$			
		Relative frequency			Bias	Relative frequency			Bias
		$< m_0$	$= m_0$	$> m_0$		$< m_0$	$= m_0$	$> m_0$	
(1) Sequential procedure based on $\widehat{D}_{j,\mathbf{K}}$									
$\widehat{m}_8$	$\gamma = 0.05$	0.080	0.877	0.043	0.077	0.003	0.956	0.041	0.133
	$\gamma = 0.01$	0.178	0.807	0.015	-0.136	0.010	0.976	0.014	0.031
$\widehat{m}_7$	$\gamma = 0.05$	0.001	0.943	0.056	0.159	0.001	0.950	0.049	0.150
	$\gamma = 0.01$	0.001	0.979	0.020	0.050	0.001	0.979	0.020	0.055
$\widehat{m}_6$	$\gamma = 0.05$	0.004	0.953	0.043	0.103	0.008	0.952	0.040	0.077
	$\gamma = 0.01$	0.334	0.657	0.009	-0.650	0.336	0.655	0.009	-0.653
$\max\{\widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.000	0.903	0.097	0.313	0.000	0.910	0.090	0.285
	$\gamma = 0.01$	0.000	0.965	0.035	0.093	0.000	0.967	0.033	0.098
$\max\{\widehat{m}_6, \widehat{m}_7\}$	$\gamma = 0.05$	0.000	0.903	0.097	0.265	0.000	0.913	0.087	0.240
	$\gamma = 0.01$	0.000	0.971	0.029	0.068	0.000	0.971	0.029	0.075
$\max\{\widehat{m}_6, \widehat{m}_7, \widehat{m}_8\}$	$\gamma = 0.05$	0.000	0.863	0.137	0.416	0.000	0.875	0.125	0.370
	$\gamma = 0.01$	0.000	0.956	0.044	0.110	0.000	0.958	0.042	0.116
(2) Wang's (1995,1999) sequential procedure based on $\sup_{k\in\mathbf{K}} \left\lvert \mathbb{W}_{j,k}^{\mathbf{Y}} \right\rvert$									
$\widehat{m}_8^*$	$\gamma = 0.05$	0.000	0.750	0.250	0.301	0.000	0.780	0.220	0.255
	$\gamma = 0.01$	0.000	0.764	0.236	0.283	0.000	0.792	0.208	0.240
	$\gamma \to 0$	0.000	0.810	0.190	0.222	0.000	0.839	0.161	0.180
$\widehat{m}_7^*$	$\gamma = 0.05$	0.000	0.730	0.270	0.341	0.000	0.713	0.287	0.367
	$\gamma = 0.01$	0.000	0.743	0.257	0.322	0.000	0.724	0.276	0.347
	$\gamma \to 0$	0.000	0.894	0.106	0.118	0.000	0.897	0.103	0.114
$\widehat{m}_6^*$	$\gamma = 0.05$	0.000	0.675	0.325	0.432	0.000	0.639	0.361	0.493
	$\gamma = 0.01$	0.000	0.688	0.312	0.410	0.000	0.656	0.344	0.466
	$\gamma \to 0$	0.000	0.937	0.063	0.069	0.000	0.936	0.064	0.072
$\max\{\widehat{m}_7^*, \widehat{m}_8^*\}$	$\gamma = 0.05$	0.000	0.548	0.452	0.571	0.000	0.557	0.443	0.556
	$\gamma = 0.01$	0.000	0.568	0.432	0.540	0.000	0.574	0.426	0.528
	$\gamma \to 0$	0.000	0.724	0.276	0.319	0.000	0.753	0.247	0.278
$\max\{\widehat{m}_6^*, \widehat{m}_7^*\}$	$\gamma = 0.05$	0.000	0.496	0.504	0.676	0.000	0.459	0.541	0.745
	$\gamma = 0.01$	0.000	0.513	0.487	0.645	0.000	0.476	0.524	0.711
	$\gamma \to 0$	0.000	0.838	0.162	0.179	0.000	0.841	0.159	0.178
$\max\{\widehat{m}_6^*, \widehat{m}_7^*, \widehat{m}_8^*\}$	$\gamma = 0.05$	0.000	0.372	0.628	0.846	0.000	0.358	0.642	0.876
	$\gamma = 0.01$	0.000	0.391	0.609	0.809	0.000	0.376	0.624	0.838
	$\gamma \to 0$	0.000	0.678	0.322	0.370	0.000	0.704	0.296	0.334

[45] Porter, J., & Yu, P. (2015). Regression discontinuity designs with unknown discontinuity points: Testing and estimation. Journal of Econometrics, 189(1), 132-147.

[46] Qiu, P., & Yandell, B. (1998). Local polynomial jump-detection algorithm in nonparametric regression. Technometrics, 40(2), 141-152.

[47] Raimondo, M. (1998). Minimax estimation of sharp change points. The Annals of Statistics, 1379-1397.

[48] Rootzén, H. (1983). The rate of convergence of extremes of stationary normal sequences. Advances in Applied Probability, 54-80.

[49] Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. Journal of the American Statistical association, 88(424), 1273-1283.

[50] Serroukh, A., Walden, A. T., & Percival, D. B. (2000). Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. Journal of the American Statistical Association, 95(449), 184-196.

[51] Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Queen's Economics Department. Working Paper, 1101.

TABLE 5
Sequential procedures to estimate the number of jumps for the Dow-Jones Industrial Average

H_0	H_1	Test statistic
Test based on sequential procedure for $\widehat{m}_7$		
		$\widehat{D}_{7,\mathbf{K}}$
$m_0 = 0$	$m_0 > 0$	4.1204
$m_0 = 1$	$m_0 > 1$	3.8619
$m_0 = 2$	$m_0 > 2$	2.2824
Critical value		$C_{0.05}$ =1.9600
		$C_{0.01}$ =2.5758
Test based on Wang's (1995,1999) sequential procedure for $\widehat{m}_7^*$		
		$\sup_{k\in\mathbf{K}} \left\lvert \mathbb{W}_{7,k}^{\mathbf{Y}} \right\rvert$
$m_0 = 0$	$m_0 > 0$	0.003481
$m_0 = 1$	$m_0 > 1$	0.002688
$m_0 = 2$	$m_0 > 2$	0.001695
$m_0 = 3$	$m_0 > 3$	0.000940
Critical value		$C_{0.05}^*$ =0.001174
		$C_{0.01}^*$ =0.001181
		$\lim_{\gamma\to 0} C_\gamma^*$ =0.001308

TABLE 6
Sequential procedures to estimate the number of jumps for the water level of the Nile River

H_0	H_1	Test statistic
Test based on sequential procedure for $\widehat{m}_7$		
		$\widehat{D}_{7,\mathbf{K}}$
$m_0 = 0$	$m_0 > 0$	4.4520
$m_0 = 1$	$m_0 > 1$	3.6423
$m_0 = 2$	$m_0 > 2$	2.9267
$m_0 = 3$	$m_0 > 3$	2.3435
Critical value		$C_{0.05}$ =1.9600
		$C_{0.01}$ =2.5758
Test based on Wang's (1995,1999) sequential procedure for $\widehat{m}_7^*$		
		$\sup_{k\in\mathbf{K}} \left\lvert \mathbb{W}_{7,k}^{\mathbf{Y}} \right\rvert$
$m_0 = 0$	$m_0 > 0$	11.7006
$m_0 = 1$	$m_0 > 1$	11.0840
$m_0 = 2$	$m_0 > 2$	10.2247
$m_0 = 3$	$m_0 > 3$	9.6761
$m_0 = 4$	$m_0 > 4$	7.1524
Critical value		$C_{0.05}^*$ =7.0303
		$C_{0.01}^*$ =7.0656
		$\lim_{\gamma\to 0} C_\gamma^*$ =7.2975

[52] Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. The Annals of Statistics, 26(4), 1356-1378.

[53] Van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge University Press.

[54] Wang, Y. (1995). Jump and sharp cusp detection by wavelets. Biometrika 82, 385-397.

[55] Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. The Annals of Statistics, 24(2), 466-484.

[56] Wang, Y. (1999). Change-points via wavelets for indirect data. Statistica Sinica, 103-117.

[57] Wu, J. S. and Chu, C. K. (1993), Kernel-type estimators of jump points and values of a regression function, The Annals of Statistics, 21, 1545-1566.

[58] Wu, W. B., & Zhao, Z. (2007). Inference of trends in time series. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(3), 391-410.

SUPPLEMENTARY MATERIAL

The supplementary materials include the following sections on: (S1) Two types of discrete wavelet transformation; (S2) Proof of Theorem 1; (S3): Proof of Theorem 2; (S4) Proof of Proposition 3; and (S5) Lemmas and their proofs.

S1: Two types of DWTs

Define two types of Discrete Wavelet Transformation (DWT) at scale $j \in \mathbb{Z}$ and location $k \in \mathbb{Z}$ for either discrete time series $\{X_t, t \in \mathbb{Z}\}$ or process $\{X(t)\}_{t\in\mathbb{R}}$ as

$$\mathbb{W}^{\mathbf{Y}}_{j,k} \equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)Y_i,$$

$$W^{\mathbf{Y}}_{j,k} \equiv \int_0^1 \psi_{j,k}(x)\,Y(dx)$$

and

$$\mathbb{W}^{\mathbf{C}}_{j,k} \equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)f_C\left(\frac{i}{n}\right),$$

$$W^{\mathbf{C}}_{j,k} \equiv \int_0^1 \psi_{j,k}(x)\,f_C(dx)$$

and

$$\begin{aligned}\mathbb{W}^{\mathbf{J}}_{j,k} &\equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)f_J\left(\frac{i}{n}\right)\\ &= \sum_{l=1}^{m_0} d_l \mathbb{W}^{\mathbf{I}_l}_{j,k},\\ W^{\mathbf{J}}_{j,k} &\equiv \int_0^1 \psi_{j,k}(x)\,f_J(dx)\\ &= \sum_{l=1}^{m_0} d_l W^{\mathbf{I}_l}_{j,k}\end{aligned}$$

and

$$\mathbb{W}^{\mathbf{B}_H}_{j,k} \equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)\varepsilon_i,$$

$$W^{\mathbf{B}_H}_{j,k} \equiv \int_0^1 \psi_{j,k}(x)\,B_H(dx)$$

where

$$\begin{aligned}\psi_{j,k}(x) &\equiv 2^{j/2}\psi\left(k - 2^j x\right),\\ \mathbb{W}^{\mathbf{I}_l}_{j,k} &\equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)I\{\frac{i}{n} > \lambda_l\},\\ W^{\mathbf{I}_l}_{j,k} &\equiv \int_0^1 \psi_{j,k}(x)\,I\{x > \lambda_l\}dx.\end{aligned}$$

Here, when $H = 0.5$, B_H is the standard Brownian motion B. When $0.5 < H < 1$, B_H is a fractional Gaussian noise that is the formal derivative of a standard fractional Brownian motion given by

$$B_H(x) \equiv \frac{1}{\Gamma(H+1/2)} \int_0^x (x-u)^{H-1/2} B(du).$$

Recall the support of wavelet function $\psi(t)$ is $[0, T]$ with $T < \infty$, then (i) The role of scale j: The support of $\psi_{j,k}(t)$ is given by $[2^{-j}k, 2^{-j}(k+T)]$ so that the finer scale, namely increasing the scale j, shrinks the support of $\psi_{j,k}(t)$. (ii) The role of location k: Increasing the location index k to k' shifts the support of $\psi_{j,k}(t)$ from $[2^{-j}k, 2^{-j}(k+T)]$ to $[2^{-j}k', 2^{-j}(k'+T)]$.

S2: Proof of Theorem 1

PROOF. (i) Under H_0, from Lemma (7.2) and (7.4), we have

$$\begin{aligned}
\widehat{\sigma}^2_{j,\mathbf{K}} &\equiv \frac{\sum_{k=1}^{2^j} (\mathbb{W}^{\mathbf{Y}}_{j,k})^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j} (\mathbb{W}^{\mathbf{C}}_{j,k} + \mathbb{W}^{\mathbf{B}_H}_{j,k})^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j} (W^{\mathbf{C}}_{j,k} + n^{H-1} W^{\mathbf{B}_H}_{j,k} + O\left(\frac{2^{j/2}}{n}\right) + O_p\left[\sqrt{\log n}\left(\frac{2^j}{n}\right)^{1-H} n^{-1/2}\right])^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j} (W^{\mathbf{C}}_{j,k} + n^{H-1} W^{\mathbf{B}_H}_{j,k} + O_p\left[\sqrt{\log n}\left(\frac{2^j}{n}\right)^{1-H} n^{-1/2}\right])^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j} \left[W^{\mathbf{C}}_{j,k} + n^{H-1} W^{\mathbf{B}_H}_{j,k} + o_p\left(n^{H-1} W^{\mathbf{B}_H}_{j,k}\right)\right]^2}{2^j}
\end{aligned}$$

where the third equality is due to $H \in [0.5, 1)$ and

$$\frac{\sqrt{\log n}\left(\frac{2^j}{n}\right)^{1-H} n^{-1/2}}{\frac{2^{j/2}}{n}} = \sqrt{\log n}\left(\frac{2^j}{n}\right)^{1/2-H} \to \infty;$$

while the fourth equality is because

$$\frac{n^{H-1} W^{\mathbf{B}_H}_{j,k}}{\sqrt{\log n}\left(\frac{2^j}{n}\right)^{1-H} n^{-1/2}} = \frac{1}{\sqrt{\log n}}\left(\frac{n}{2^j}\right)^{1/2} \to \infty.$$

And we have

$$\begin{aligned}
\widetilde{\sigma}^2_{j,\mathbf{K}} &\equiv \left[med_{k\in\{1,\cdots,2^j\}} \left| \frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{0.6745} \right| \right]^2 \\
&= \left[med_{k\in\{1,\cdots,2^j\}} \left| \frac{W^{\mathbf{C}}_{j,k} + n^{H-1} W^{\mathbf{B}_H}_{j,k} + O_p\left[\sqrt{\log n}\left(\frac{2^j}{n}\right)^{1-H} n^{-1/2}\right]}{0.6745} \right| \right]^2
\end{aligned}$$

$$= \left[med_{k\in\{1,\cdots,2^j\}} \left| \frac{W_{j,k}^{\mathbf{C}} + n^{H-1} W_{j,k}^{\mathbf{B}_H} + o_p\left(n^{H-1} W_{j,k}^{\mathbf{B}_H}\right)}{0.6745} \right| \right]^2 .$$

Since $\left\{W_{j,k}^{\mathbf{B}_H}\right\}_{k=1}^{2^j}$ is short-range dependent according to Wang (1996), we could show that

$$\begin{aligned} & Var\left[\frac{\sum_{k=1}^{2^j} n^{2H-2}\left(W_{j,k}^{\mathbf{B}_H}\right)^2}{2^j}\right] \\ &= O\left\{\frac{n^{4H-4} Var\left[\left(W_{j,1}^{\mathbf{B}_H}\right)^2\right]}{2^j}\right\} \\ &= O\left\{\frac{n^{4H-4} 2\left[Var\left(W_{j,1}^{\mathbf{B}_H}\right)\right]^2}{2^j}\right\} \\ &= O\left[\frac{n^{4H-4}}{2^{(4H-1)j}}\right] \end{aligned}$$

where the second equality is from $Var\left[\left(W_{j,1}^{\mathbf{B}_H}\right)^2\right] = 2\left[Var\left(W_{j,1}^{\mathbf{B}_H}\right)\right]^2$ due to the Gaussianity of $W_{j,1}^{\mathbf{B}_H}$ (Papoulis 1984, p. 233). Similarly, $Var\left(\widetilde{\sigma}_{j,\mathbf{K}}^2\right) = O\left[\frac{n^{4H-4}}{2^{(4H-1)j}}\right]$. By Chebyshev's Inequality, we have

$$\begin{aligned} \frac{\sum_{k=1}^{2^j}\left(n^{H-1} W_{j,k}^{\mathbf{B}_H}\right)^2}{2^j} &= O_p\left[E\left(\frac{\sum_{k=1}^{2^j}\left(n^{H-1} W_{j,k}^{\mathbf{B}_H}\right)^2}{2^j}\right) + \sqrt{Var\left(\frac{\sum_{k=1}^{2^j} n^{2H-2}\left(W_{j,k}^{\mathbf{B}_H}\right)^2}{2^j}\right)}\right] \\ &= O_p\left[Var\left(n^{H-1} W_{j,k}^{\mathbf{B}_H}\right) + \frac{n^{2H-2}}{2^{(2H-1/2)j}}\right] \\ &= O_p\left[\frac{n^{2H-2}}{2^{(2H-1)j}} + \frac{n^{2H-2}}{2^{(2H-1/2)j}}\right] \\ &= O_p\left[\frac{n^{2H-2}}{2^{(2H-1)j}}\right] \end{aligned}$$

and

$$\begin{aligned} \frac{\sum_{k=1}^{2^j} W_{j,k}^{\mathbf{C}} n^{H-1} W_{j,k}^{\mathbf{B}_H}}{2^j} &= O_p\left[\sqrt{Var\left(\frac{\sum_{k=1}^{2^j} W_{j,k}^{\mathbf{C}} n^{H-1} W_{j,k}^{\mathbf{B}_H}}{2^j}\right)}\right] \\ &= O_p\left[\frac{n^{H-1}}{2^{(H+3/2)j}}\right]. \end{aligned}$$

Thus, under Assumption 3(i) where $\frac{n^{(H-1)/(H-3/2)}}{2^j} \to 0$, we have

$$\frac{\sum_{k=1}^{2^j}\left(n^{H-1}W_{j,k}^{\mathbf{B}_H}\right)^2}{\sum_{k=1}^{2^j}W_{j,k}^{\mathbf{C}}n^{H-1}W_{j,k}^{\mathbf{B}_H}} = O_p\left[\frac{n^{2H-2}2^{(H+3/2)j}}{2^{(2H-1)j}n^{H-1}}\right]$$

$$= O_p\left[\frac{n^{H-1}}{2^{(H-5/2)j}}\right] \to \infty,$$

$$\frac{\sum_{k=1}^{2^j}\left(n^{H-1}W_{j,k}^{\mathbf{B}_H}\right)^2}{\sum_{k=1}^{2^j}\left(W_{j,k}^{\mathbf{C}}\right)^2} = O_p\left[\frac{n^{2H-2}}{2^{(2H-1)j}2^{-3j}}\right]$$

$$= O_p\left[\frac{n^{H-1}}{2^{(H-2)j}}\right]^2 \to \infty,$$

$$\sup_{k\in\mathbf{K}}\left(\frac{n^{H-1}W_{j,k}^{\mathbf{B}_H}}{W_{j,k}^{\mathbf{C}}}\right) = O_p\left[\frac{n^{H-1}}{2^{(H-2)j}}\right] \to \infty.$$

Notice that under Assumption 3(i),

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}}\frac{\sum_{k=1}^{2^j}\left(W_{j,k}^{\mathbf{C}}\right)^2}{2^j} = O\left[\frac{2^{(2H-7/2)j}}{n^{2H-2}}\right] \to 0$$

and

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}}\frac{\sum_{k=1}^{2^j}W_{j,k}^{\mathbf{C}}n^{H-1}W_{j,k}^{\mathbf{B}_H}}{2^j} = O_p\left[\frac{2^{(2H-1/2)j}}{n^{2H-2}}\frac{n^{H-1}}{2^{(H+3/2)j}}\right]$$

$$= O_p\left[\frac{2^{(H-2)j}}{n^{H-1}}\right] \to 0.$$

Hence, both terms $\frac{\sum_{k=1}^{2^j}\left(W_{j,k}^{\mathbf{C}}\right)^2}{2^j}$ and $\frac{\sum_{k=1}^{2^j}W_{j,k}^{\mathbf{C}}n^{H-1}W_{j,k}^{\mathbf{B}_H}}{2^j}$ do not introduce the bias into the limiting distribution of $\widehat{\sigma}_{j,\mathbf{K}}^2$. According to Kouamo et al. (2013), where they provide asymptotic expansions for $\widehat{\sigma}_{j,\mathbf{K}}^2$ and $\widetilde{\sigma}_{j,\mathbf{K}}^2$, we can derive the following multivariate central limit theorem

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}}[]\widehat{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\widetilde{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2 \xrightarrow{d} N\left[\mathbf{0},\begin{pmatrix} v_1^* & v_{12}^* \\ v_{12}^* & v_2^*\end{pmatrix}\right]$$

where

$$v_1^* \equiv \lim_{n\to\infty}\frac{2^{(4H-1)j}}{n^{4H-4}}Var\left[\frac{\sum_{k=1}^{2^j}\left(\mathbb{W}_{j,k}^{\mathbf{B}_H}\right)^2}{2^j}\right]$$

$$v_2^* = \lim_{n\to\infty}\frac{2^{(4H-1)j}}{n^{4H-4}}Var\left\{\left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{\mathbb{W}_{j,k}^{\mathbf{B}_H}}{0.6745}\right|\right]^2\right\}$$

$$v_{12}^* \equiv \lim_{n\to\infty}\frac{2^{(4H-1)j}}{n^{4H-4}}Cov\left\{\frac{\sum_{k=1}^{2^j}\left(\mathbb{W}_{j,k}^{\mathbf{B}_H}\right)^2}{2^j},\left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{\mathbb{W}_{j,k}^{\mathbf{B}_H}}{0.6745}\right|\right]^2\right\}$$

and $\begin{pmatrix} v_1^* & v_{12}^* \\ v_{12}^* & v_2^* \end{pmatrix}$ is a 2×2 positive definite matrix. Notice that the asymptotic covariance matrix $\begin{pmatrix} v_1^* & v_{12}^* \\ v_{12}^* & v_2^* \end{pmatrix}$ can be further derived according to the influence functions of $\widehat{\sigma}^2_{j,\mathbf{K}}$ and $\widetilde{\sigma}^2_{j,\mathbf{K}}$ from Proposition 1 of Kouamo et al. (2013).[14] Therefore, we have

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}} \left(\widehat{\sigma}^2_{j,\mathbf{K}} - \widetilde{\sigma}^2_{j,\mathbf{K}}\right) \xrightarrow{d} N\left[0, (1,-1)\begin{pmatrix} v_1^* & v_{12}^* \\ v_{12}^* & v_2^* \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right].$$

(ii) Under H_1, from Lemma (7.2), (7.3) and (7.4), we have

$$\begin{aligned}
\widehat{\sigma}^2_{j,\mathbf{K}} &\equiv \frac{\sum_{k=1}^{2^j}(\mathbb{W}^{\mathbf{Y}}_{j,k})^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j}(\mathbb{W}^{\mathbf{C}}_{j,k} + \mathbb{W}^{\mathbf{J}}_{j,k} + \mathbb{W}^{\mathbf{B}_H}_{j,k})^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j}\left[W^{\mathbf{C}}_{j,k} + W^{\mathbf{J}}_{j,k} + n^{H-1}W^{\mathbf{B}_H}_{j,k} + o_p\left(n^{H-1}W^{\mathbf{B}_H}_{j,k}\right)\right]^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j}\left[W^{\mathbf{C}}_{j,k} + \sum_{l=1}^{m_0} d_l W^{\mathbf{I}_l}_{j,k} + n^{H-1}W^{\mathbf{B}_H}_{j,k} + o_p\left(n^{H-1}W^{\mathbf{B}_H}_{j,k}\right)\right]^2}{2^j} \\
&= \frac{\sum_{k=1}^{2^j}\left[W^{\mathbf{C}}_{j,k} + n^{H-1}W^{\mathbf{B}_H}_{j,k} + o_p\left(n^{H-1}W^{\mathbf{B}_H}_{j,k}\right)\right]^2}{2^j} \\
&\quad + \frac{\sum_{k=1}^{2^j}\left[\sum_{l=1}^{m_0} d_l W^{\mathbf{I}_l}_{j,k} + o_p\left(n^{H-1}W^{\mathbf{B}_H}_{j,k}\right)\right]^2}{2^j} \\
&\quad + \frac{2\sum_{k=1}^{2^j}\left\{\left[W^{\mathbf{C}}_{j,k} + n^{H-1}W^{\mathbf{B}_H}_{j,k} + o_p\left(n^{H-1}W^{\mathbf{B}_H}_{j,k}\right)\right]\left[\sum_{l=1}^{m_0} d_l W^{\mathbf{I}_l}_{j,k} + o_p\left(n^{H-1}W^{\mathbf{B}_H}_{j,k}\right)\right]\right\}}{2^j}
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\sigma}^2_{j,\mathbf{K}} &\equiv \left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{\mathbb{W}^{\mathbf{Y}}_{j,k}}{0.6745}\right|\right]^2 \\
&= \left[med_{k\in\{1,\cdots,2^j\}}\left|\frac{W^{\mathbf{C}}_{j,k} + \sum_{l=1}^{m_0} d_l W^{\mathbf{I}_l}_{j,k} + n^{H-1}W^{\mathbf{B}_H}_{j,k} + o_p\left(n^{H-1}W^{\mathbf{B}_H}_{j,k}\right)}{0.6745}\right|\right]^2.
\end{aligned}$$

Because the number of non-zero terms in $\sum_{k=1}^{2^j}\sum_{l=1}^{m_0}\left(W^{\mathbf{I}_l}_{j,k}\right)^2$ is finite and $\left(W^{\mathbf{I}_l}_{j,k}\right)^2 = O_p\left(2^{-j}\right)$ from Lemma (7.3), we have

$$\sum_{l=1}^{m_0}\left[d_l^2\sum_{k=1}^{2^j}\left(W^{\mathbf{I}_l}_{j,k}\right)^2\right]$$

[14] Here we omit this further expression, and interested readers can derive it based on Lévy-Leduc et al. (2011).

$$= \sum_{k=1}^{2^j} \left(W_{j,k}^{\mathbf{I}_l} \right)^2 \cdot \sum_{l=1}^{m_0} d_l^2$$
$$= O\left(2^{-j}\right),$$

thus

$$\frac{\sum_{k=1}^{2^j} \left[\sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l} \right]^2}{2^j} = O\left(2^{-2j}\right).$$

Similarly, since the number of non-zero terms in $\sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}$ is finite and $W_{j,k}^{\mathbf{I}_l} = O_p\left(2^{-j/2}\right)$ from Lemma (7.3), then we have

$$\frac{\sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l} \right)}{2^j} = O_p\left(\frac{n^{H-1} W_{j,k}^{\mathbf{B}_H} W_{j,k}^{\mathbf{I}_l}}{2^j} \right)$$
$$= O_p\left[\sqrt{Var\left(\frac{n^{H-1} W_{j,k}^{\mathbf{B}_H} W_{j,k}^{\mathbf{I}_l}}{2^j} \right)} \right]$$
$$= O_p\left[\frac{n^{H-1}}{2^{(H+1)j}} \right].$$

Notice that

$$\frac{\sum_{k=1}^{2^j} \left[\sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l} \right]^2}{\sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l} \right)} = O_p\left(\frac{2^{-2j} 2^{(H+1)j}}{n^{H-1}} \right)$$
$$= O_p\left(\frac{2^{(H-1)j}}{n^{H-1}} \right) \to \infty.$$

Therefore

$$\begin{aligned} \widehat{\sigma}_{j,\mathbf{K}}^2 = & \frac{\sum_{k=1}^{2^j} n^{2H-2} \left(W_{j,k}^{\mathbf{B}_H} \right)^2}{2^j} \\ & + \frac{2 \sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l} \right)}{2^j} \\ & + \frac{\sum_{k=1}^{2^j} \left[\sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l} \right]^2}{2^j} \\ & + o_p\left[\frac{2 \sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l} \right)}{2^j} \right]. \end{aligned}$$

Notice that the second and third terms are introduced because of the presence of jumps under H_1. Thus under H_1

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}} \left(\widehat{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2 \right)$$

$$= \frac{2^{(2H-1/2)j}}{n^{2H-2}} \left[\begin{array}{c} \frac{\sum_{k=1}^{2^j} n^{2H-2} \left(W_{j,k}^{\mathbf{B}_H}\right)^2}{2^j} - \sigma_j^2 \\ + \frac{2 \sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right)}{2^j} + \frac{\sum_{k=1}^{2^j} \left[\sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right]^2}{2^j} \\ + o_p \left[\frac{2 \sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right)}{2^j} \right] \end{array} \right]$$

$$= N(0, v_1^*) + \frac{2^{(2H-1/2)j}}{n^{2H-2}} \frac{2 \sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right)}{2^j} + \frac{2^{(2H-1/2)j}}{n^{2H-2}} \frac{\sum_{k=1}^{2^j} \left[\sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right]^2}{2^j} + o_p \left[\frac{2^{(2H-1/2)j}}{n^{2H-2}} \frac{2 \sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right)}{2^j} \right].$$

On the other hand,

$$\frac{2^{(2H-1/2)j}}{n^{2H-2}} \left(\widetilde{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\right) \to N(0, v_2^*)$$

regardless of being under H_0 or H_1. Recall

$$\widehat{D}_{j,\mathbf{K}} = \frac{\widehat{\sigma}_{j,\mathbf{K}}^2 - \widetilde{\sigma}_{j,\mathbf{K}}^2}{\sqrt{\widehat{\omega}}} = \frac{\left(\widehat{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\right) - \left(\widetilde{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\right)}{\sqrt{\widehat{\omega}}} = \frac{\frac{2^{(2H-1/2)j}}{n^{2H-2}} \left(\widehat{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\right) - \frac{2^{(2H-1/2)j}}{n^{2H-2}} \left(\widetilde{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\right)}{\sqrt{\frac{2^{(4H-1)j}}{n^{4H-4}} \omega}} \frac{\sqrt{\omega}}{\sqrt{\widehat{\omega}}} = \frac{\frac{2^{(2H-1/2)j}}{n^{2H-2}} \left(\widehat{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\right) - \frac{2^{(2H-1/2)j}}{n^{2H-2}} \left(\widetilde{\sigma}_{j,\mathbf{K}}^2 - \sigma_j^2\right)}{\sqrt{(1,-1) \begin{pmatrix} v_1^* & v_{12}^* \\ v_{12}^* & v_2^* \end{pmatrix} \binom{1}{-1}}} \frac{\sqrt{\omega}}{\sqrt{\widehat{\omega}}}$$

where

$$\frac{2^{(4H-1)j}}{n^{4H-4}} \omega = (1,-1) \begin{pmatrix} v_1^* & v_{12}^* \\ v_{12}^* & v_2^* \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

by construction. Since $\widehat{\omega}$ is a consistent estimate of ω, we have

$$\widehat{D}_{j,\mathbf{K}} = O_p \left[\frac{2^{(2H-1/2)j}}{n^{2H-2}} \frac{2 \sum_{k=1}^{2^j} \left(n^{H-1} W_{j,k}^{\mathbf{B}_H} \sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right)}{2^j} \right] + O_p \left[\frac{2^{(2H-1/2)j}}{n^{2H-2}} \frac{\sum_{k=1}^{2^j} \left(\sum_{l=1}^{m_0} d_l W_{j,k}^{\mathbf{I}_l}\right)^2}{2^j} \right]$$

$$
\begin{aligned}
&+O_p(1) \\
&= O_p\left[\frac{2^{(2H-1/2)j}}{n^{2H-2}}\frac{n^{H-1}}{2^{(H+1)j}}\right] + O\left[\frac{2^{(2H-1/2)j}}{n^{2H-2}}2^{-2j}\right] + O_p(1) \\
&= O_p\left[\frac{2^{(H-3/2)j}}{n^{H-1}}\right] + O\left[\frac{2^{(2H-5/2)j}}{n^{2H-2}}\right] + O_p(1) \\
&= o_p(1) + \infty + O_p(1) \\
&\to \infty
\end{aligned}
$$

where the fourth equality comes from Assumption 3 where $\frac{n^{(H-1)/(H-3/2)}}{2^j} \to 0$, while $\frac{n^{(2H-2)/(2H-5/2)}}{2^j} \to \infty$. □

S3: Proof of Theorem 2

PROOF. Following Theorem 1, we can easily show that with probability $1-\gamma$,

$$
|D_{j,\mathbf{K}}| \geq \Phi^{-1}\left(1-\frac{\gamma}{2}\right),
$$

when there are jumps at $k/2^j$ with $k \in \left\{1,\cdots,2^j\right\}$. From the definition of $\widehat{m}$ and $\widehat{k}_1,\cdots,\widehat{k}_{\widehat{m}_0}$, we can see that, with probability tending to $1-\gamma$,

$$
\widehat{m} = m_0,
$$

$$
\frac{\widehat{k}_l}{2^j} - \lambda_l = O_p\left(2^{-j}\right)
$$

for $l = 1,\cdots,\widehat{m}$. Finally, the theorem is proved by letting $\gamma \to 0$. □

S4: Proof of Proposition 3

PROOF. Under Assumption 4, it is straightforward to prove that $\frac{2^{(4H-1)j}}{n^{4H-4}}(\widehat{\omega}-\omega) = o_p(1)$ under H_0 following Andrews (1991). And under H_1, since the outliers from jumps are truncated from $\widehat{\omega}$, we still have $\frac{2^{(4H-1)j}}{n^{4H-4}}(\widehat{\omega}-\omega) = o_p(1)$ based on Assumption 4. □

S5: Lemmas and their proofs

The following lemmas provide both asymptotic decay rates and discretization bias between W and $\mathbb{W}$. Recall that $\mathbb{W}$ is DWT for the discrete time series $\{X_t, t \in \mathbb{Z}\}$, while W is DWT for the process $\{X(t)\}_{t\in\mathbb{R}}$. The reasons for proving the discretization bias between W and $\mathbb{W}$ are: (i) most properties of wavelet transformation (e.g., decorrelation and spatial adaptivity) are defined on W instead of on $\mathbb{W}$; (ii) the convergence rate is easier to obtain from W than from $\mathbb{W}$.

LEMMA 7.1. *Suppose that the first derivatives of g and ψ exist except for a small number of points. Moreover, assume that g' and ψ' (where they exist) are piecewise, continuous and bounded. Then for $j \in \mathbb{Z}$ and $k \in \left\{1,\cdots,2^j\right\} = \mathbf{K}$, we have*

$$
\mathbb{W}^{\mathbf{g}}_{j,k} = W^{\mathbf{g}}_{j,k} + O\left(\frac{2^{j/2}}{n}\right)
$$

where

$$\mathbb{W}^{\mathbf{g}}_{j,k} \equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)g\left(\frac{i}{n}\right)$$
$$W^{\mathbf{g}}_{j,k} \equiv \int_0^1 2^{j/2}\psi\left(k-2^jx\right)g\left(dx\right).$$

PROOF. Recall the rectangle rule, which is

$$\int_c^d \omega(x)dx = \frac{d-c}{n}\sum_{i=0}^{n-1}\omega(c+i\frac{d-c}{n}) + O\left(\sum_{i=0}^{n-1}\sup_{x\in C_i}\left|\omega'(x)\right|\cdot\frac{(d-c)^2}{n^2}\right)$$

with $C_i = \left[c+i\frac{d-c}{n}, c+(i+1)\frac{d-c}{n}\right]$. Noting that the support of $\psi\left(k-2^j\frac{i}{n}\right)$ (as a function of i) is $\left[kn2^{-j}, (\mathtt{T}+k)\,n2^{-j}\right]$, we obtain

$$\mathbb{W}^{\mathbf{g}}_{j,k} = 2^{j/2}\sum_{i=c_1(k)}^{c_2(k)}\frac{1}{n}\psi\left(k-2^j\frac{i}{n}\right)g\left(\frac{i}{n}\right)$$

with

$$c_1\left(k\right) \equiv kn2^{-j}$$
$$c_2\left(k\right) \equiv \left(\mathtt{T}+k\right)n2^{-j}.$$

Thus, the number of non-zero terms in the sum is $\mathtt{T}n2^{-j}$. This outcome, together with the rectangle rule for $\omega(i/n) \equiv \psi\left(k-2^j\frac{i}{n}\right)g\left(\frac{i}{n}\right)$, implies that

$$\mathbb{W}^{\mathbf{g}}_{j,k} = W^{\mathbf{g}}_{j,k} + O(\frac{2^{j/2}}{n}).$$

Note that here the factor 2^j from the derivative of $\psi\left(k-2^jx\right)$ is compensated for by the fact that the number of non-zero terms in the sum is proportional to 2^{-j}. Now, assume, more generally, that g' and ψ' exist except for a few points. The result then follows by a piecewise application of the rectangle rule.

□

Recall

$$\mathbb{W}^{\mathbf{C}}_{j,k} \equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)f_C\left(\frac{i}{n}\right),$$
$$W^{\mathbf{C}}_{j,k} \equiv \int_0^1 2^{j/2}\psi\left(k-2^jx\right)f_C\left(dx\right).$$

LEMMA 7.2. *For $j\in\mathbb{Z}$ and $k\in\left\{1,\cdots,2^j\right\}=\mathbf{K}$, we have:*
(i)

$$W^{\mathbf{C}}_{j,k} = O\left(2^{-3j/2}\right);$$

(ii)

$$\mathbb{W}^{\mathbf{C}}_{j,k} - W^{\mathbf{C}}_{j,k} = O\left(\frac{2^{j/2}}{n}\right).$$

PROOF. (i) According to Theorem 2.9.1 - 2.9.4 of Daubechies (1992), the asymptotic decay of $W_{j,k}^{\mathbf{C}}$ is the order $O\left(2^{-3j/2}\right)$. (ii) It is from Lemma (7.1).
Recall

$$\begin{aligned}\mathbb{W}_{j,k}^{\mathbf{J}} &\equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)f_J\left(\frac{i}{n}\right)\\ &= \sum_{l=1}^{m_0} d_l \cdot \mathbb{W}_{j,k}^{\mathbf{I}_l},\\ W_{j,k}^{\mathbf{J}} &\equiv \int_0^1 2^{j/2}\psi\left(k-2^j x\right)f_J\left(dx\right)\\ &= \sum_{l=1}^{m_0} d_l \cdot W_{j,k}^{\mathbf{I}_l}\end{aligned}$$

where

$$\begin{aligned}\mathbb{W}_{j,k}^{\mathbf{I}_l} &\equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)I\{\frac{i}{n}>\lambda_l\},\\ W_{j,k}^{\mathbf{I}_l} &\equiv \int_0^1 2^{j/2}\psi\left(k-2^j x\right)I\{x>\lambda_l\}dx.\end{aligned}$$

with $l \in \{1,..,m_0\}$. Let the support of the wavelet function $\psi(t)$ is $[0,T]$ with $T<\infty$. □

LEMMA 7.3. *For $j\in\mathbb{Z}$, $k\in\left\{1,\cdots,2^j\right\}=\mathbf{K}$ and $l\in\{1,..,m_0\}$, we have:*
(i)

$$W_{j,k}^{\mathbf{I}_l} = \{.O\left(2^{-j/2}\right),\ \textit{for}\ k\in\left[2^j\lambda_l, 2^j\lambda_l+\mathtt{T}\right]\cap\left\{1,\cdots,2^j\right\}0, \textit{otherwise};$$

(ii)

$$\mathbb{W}_{j,k}^{\mathbf{I}_l} - W_{j,k}^{\mathbf{I}_l} = O\left(\frac{2^{j/2}}{n}\right).$$

PROOF. (i) Using the definition $\int_0^1 2^{j/2}\psi\left(k-2^j x\right)I\{x>\lambda_l\}dx$ and applying the change of variable, we have

$$\begin{aligned}&\frac{1}{2^{j/2}}\int_0^T \psi\left(v\right)I\{k-v>2^j\lambda_l\}dv\\ &\qquad = \{.O\left(2^{-j/2}\right),\ \text{for}\ k\in\left[2^j\lambda_l,2^j\lambda_l+T\right]\cap\left\{1,\cdots,2^j\right\}0,\text{otherwise}.\end{aligned}$$

(ii) This result follows from Lemma (7.1) or Lemma D.8 of Chen and Fan (2019). □

Recall

$$\begin{aligned}\mathbb{W}_{j,k}^{\mathbf{B}_H} &\equiv \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{i}{n}\right)\varepsilon_i,\\ W_{j,k}^{\mathbf{B}_H} &\equiv \int_0^1 2^{j/2}\psi\left(k-2^j x\right)B_H\left(dx\right).\end{aligned}$$

LEMMA 7.4. *For $j \in \mathbb{Z}$ and $k \in \left\{1, \cdots, 2^j\right\} = \mathbf{K}$, we have:*
(i)

$$W_{j,k}^{\mathbf{B}_H} = O_P\left[2^{-(H-1/2)j}\right]$$
$$\sup_{k\in\mathbf{K}} W_{j,k}^{\mathbf{B}_H} = O_P\left[2^{-(H-1/2)j}\left|\log 2^j\right|^{1/2}\right];$$

(ii)

$$\max_{k\in\mathbf{K}}\left|n^{H-1/2}W_{j,k}^{\mathbf{B}_H} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}_H}\right| = O_p\left[\sqrt{\log n}\left(\frac{2^j}{n}\right)^{1-H}\right].$$

PROOF. (i) When $H = 0.5$, the asymptotic decay of $W_{j,k}^{\mathbf{B}_H}$ is the order of $O_p(1)$ and $\sup_{k\in\mathbf{K}} W_{j,k}^{\mathbf{B}_H} = O_p\left(\left|\log 2^j\right|^{1/2}\right)$ based on Wang (1995). For $0.5 < H < 1$, the results follow Wang (1996). (ii) Our proof depends on whether or not $H = 0.5$. When $H = 0.5$, the integration with respect to the Brownian motion B, so that we need to apply the Itô calculus in the derivation. On the other hand, when $0.5 < H < 1$, B_H, is not semimartingale, so that the Itô calculus can't be applied, and we refer to Mémin, Mishura and Valkeila (2001) to bound the moments of Wiener integrals with respect to fractional Brownian motions.[15]
Case 1 of $H = 0.5$ where $B = B_{1/2}$, then

$$W_{j,k}^{\mathbf{B}} = \int_0^{2^j} 2^{j/2}\psi\left(k - 2^j x\right) dB\left(2^j x\right)\cdot\frac{1}{2^{j/2}}$$
$$= \int_0^{2^j} \psi\left(k - t\right) dB\left(t\right)$$

where the first equality comes from the scaling invariance of the Brownian motion. Recall

$$\sqrt{n}\mathbb{W}_{j,k}^{\mathbf{B}} = \frac{\sum_{i=1}^n 2^{j/2}\psi\left(k - \frac{i}{n}2^j\right)\varepsilon_i}{\sqrt{n}}$$
$$= \int_0^{2^j}\psi\left(k - \frac{\left\lfloor 1 + tn/2^j\right\rfloor}{n/2^j}\right) d\mathbb{B}\left(t\right)$$

where the second equality occurs because $\varepsilon_i := B\left(i\right) - B\left(i-1\right)$. Notice that both $W_{j,k}^{\mathbf{B}}$ and $\sqrt{n}\mathbb{W}_{j,k}^{\mathbf{B}}$ are stationary Gaussian processes. In order to show

$$\max_{k\in\mathbb{R}}\left|W_{j,k}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}}\right| = O_p\left(\sqrt{\frac{\log n}{n/2^j}}\right),$$

we apply the chaining argument of Cranston et al. (2000). Notice that

$$E\left[\int_0^{2^j}\psi\left(k-t\right)dB\left(t\right) - \int_0^{2^j}\psi\left(k - \frac{\left\lfloor 1 + tn/2^j\right\rfloor}{n/2^j}\right)dB\left(t\right)\right]^2$$
$$\le E\left[\int_0^{2^j}\sup_{v,v'\in\left[k-t-n2^{-j},k-t+n2^{-j}\right]}\left|\psi\left(v\right) - \psi\left(v'\right)\right| dB\left(t\right)\right]^2$$

[15] It might be possible to integrate both cases of $H = 1/2$ and $H > 1/2$ by using the harmonizable representations of wavelet coefficients; see Ayache and Bertrand (2011) but their results are only pointwise.

$$= \int_0^{2^j} \sup_{v,v' \in [k-t-2^j n^{-1}, k-t+2^j n^{-1}]} \left|\psi(v) - \psi(v')\right|^2 dt$$

$$= \frac{C}{n/2^j}$$

where the first equality is due to the Itô's Isometry. Similarly, if

$$\left|k - k'\right| \le \frac{1}{2},$$

we have

$$E\left[W_{j,k}^{\mathbf{B}} - W_{j,k'}^{\mathbf{B}}\right]^2 \le C\left|k - k'\right|$$

$$E\left[n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k'}^{\mathbf{B}}\right]^2 \le C\left|k - k'\right|,$$

so that

$$E\left[W_{j,k}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}} - \left(W_{j,k'}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k'}^{\mathbf{B}}\right)\right]^2 \le 4C\left|k - k'\right|.$$

Let $\epsilon := 2^{-1}, \epsilon_l := (l+3)^{-2}, \delta_l := \left(2^l n^2\right)^{-1}$ and $\mathcal{X}_l := \{k\delta_l, k \in \mathbb{Z}\} \cap \mathbb{R}, l \ge 0$. Then the cardinality $|\mathcal{X}_l| \le 2^j/\delta_l$ and $\epsilon + \sum_{l=0}^{\infty} \epsilon_l < 1$. Let $\lambda = 8\sqrt{C}$. By Lemma 4.1 in Cranston et al. (2000), we have

$$\Pr\left(\max_{k\in\mathbb{R}}\left|W_{j,k}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}}\right| > \lambda\sqrt{\frac{\log n}{n/2^j}}\right)$$

$$\le \Pr\left\{\left|W_{j,k_0}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k_0}^{\mathbf{B}}\right| > \lambda\epsilon\sqrt{\frac{\log n}{n/2^j}}\right\}$$

$$+ \sum_{l=0}^{\infty} |\mathcal{X}_l| \sup_{|k-k'|\le\delta_l} \Pr\left\{\left|W_{j,k}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}} - \left(W_{j,k'}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k'}^{\mathbf{B}}\right)\right| > \lambda\epsilon_l\sqrt{\frac{\log n}{n/2^j}}\right\}$$

$$\le 2\left[1 - \Phi\left(\lambda\epsilon\sqrt{\frac{\log n}{n/2^j}}\sqrt{\frac{n/2^j}{C}}\right)\right] + \sum_{l=0}^{\infty} 2\frac{2^j}{\delta_l}\left[1 - \Phi\left(\lambda\epsilon_l\sqrt{\frac{\log n}{n/2^j}}\sqrt{\frac{1}{C\delta_l}}\right)\right].$$

Since $1 - \Phi(t) \sim \phi(t)/t$ as $t \to \infty$, we can show that

$$\Pr\left(\max_{k\in\mathbb{R}}\left|W_{j,k}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}}\right| > \lambda\sqrt{\frac{\log n}{n/2^j}}\right) = O\left(n^{-2}\right).$$

Then

$$\max_{k\in\mathbf{K}}\left|W_{j,k}^{\mathbf{B}} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}}\right| = O_p\left(\sqrt{\frac{\log n}{n/2^j}}\right).$$

Case 2 of $0.5 < H < 1$, then

$$W_{j,k}^{\mathbf{B}_H} = \int_0^{2^j} 2^{j/2}\psi\left(k - 2^j u\right) dB_H\left(2^j u\right) \cdot \frac{1}{2^{Hj}}$$

$$= \int_0^{2^j} 2^{(1/2-H)j}\psi(k-t)\, dB_H(t)$$

where the first equality comes from the scaling invariance of the fractional Brownian motion. Recall

$$n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}_H} = \frac{\sum_{i=1}^{n} 2^{j/2}\psi\left(k - \frac{i}{n}2^j\right)\varepsilon_i}{\sqrt{n}}$$
$$= \int_0^{2^j} \left(\frac{2^j}{n}\right)^{1/2-H} \psi\left(k - \frac{\lfloor 1 + tn/2^j \rfloor}{n/2^j}\right) dB_H\left(t\right)$$

where the second equality occurs because $\varepsilon_i \equiv B_H(i) - B_H(i-1)$ and the scaling invariance of the fractional Brownian motion. Notice that both $\int_0^{2^j} \psi(k-t)\, dB_H(t)$ and $\int_0^{2^j} \psi\left(k - \frac{\lfloor 1+tn/2^j \rfloor}{n/2^j}\right) dB_H(t)$ are stationary Gaussian processes. Now we have

$$E\left[\int_0^{2^j} \psi\left(k-t\right) dB_H\left(t\right) - \int_0^{2^j} \psi\left(k - \frac{\lfloor 1 + tn/2^j \rfloor}{n/2^j}\right) dB_H\left(t\right)\right]^2$$
$$\leq E\left[\int_0^{2^j} \sup_{v,v' \in [k-t-n2^{-j},k-t+n2^{-j}]} \left|\psi\left(v\right) - \psi\left(v'\right)\right| dB_H\left(t\right)\right]^2$$
$$\leq C\left(\int_0^{2^j} \sup_{v,v' \in [k-t-2^j n^{-1},k-t+2^j n^{-1}]} \left|\psi\left(v\right) - \psi\left(v'\right)\right|^{1/H} dt\right)^{2H}$$
$$= \frac{C}{n/2^j}$$

where the second inequality is due to Mémin et al. (2001). Analogously to the proof of Case 1 of $H = 0.5$ (or the chaining argument of Lemma 4.1 in Cranston et al., 2000) we have

$$\max_{k \in \mathbf{K}} \left|\int_0^{2^j} \psi\left(k-t\right) dB_H\left(t\right) - \int_0^{2^j} \psi\left(k - \frac{\lfloor 1 + tn/2^j \rfloor}{n/2^j}\right) dB_H\left(t\right)\right| = O_p\left(\sqrt{\frac{\log n}{n/2^j}}\right),$$

in which the right hand side is independent of H. In the end, since

$$n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}_H} = \left(\frac{2^j}{n}\right)^{1/2-H} \int_0^{2^j} \psi\left(k - \frac{\lfloor 1 + tn/2^j \rfloor}{n/2^j}\right) dB_H\left(t\right)$$

and

$$W_{j,k}^{\mathbf{B}_H} = 2^{(1/2-H)j} \int_0^{2^j} \psi\left(k-t\right) dB_H\left(t\right),$$

then we have

$$\max_{k \in \mathbf{K}} \left|n^{H-1/2} W_{j,k}^{\mathbf{B}_H} - n^{1/2}\mathbb{W}_{j,k}^{\mathbf{B}_H}\right| = O_p\left[\sqrt{\log n}\left(\frac{2^j}{n}\right)^{1-H}\right].$$

□