

Bootstrap testing for discontinuities under long-range dependence

Jan Beran*, Yevgen Shumeyko

Department of Mathematics and Statistics, University of Konstanz, Germany

ARTICLE INFO

Article history:

Received 3 February 2011

Available online 19 October 2011

AMS subject classifications:

62G08

62G09

62M10

62M07

Keywords:

Long-range dependence

Bootstrap

Nonparametric regression

Wavelet

Thresholding

Test for discontinuity

ABSTRACT

We consider testing for discontinuities in a trend function when the residual process exhibits long memory. Using a wavelet decomposition of the estimated trend function into a low-resolution and a high-resolution component, a test statistic is proposed based on blockwise resampling of estimated residual variances. Asymptotic validity of the test is derived. A simulation study illustrates finite sample properties.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Consider time series data of the form

$$Y_i = g(t_i) + \xi_i, \quad (i = 1, 2, \dots, n) \quad (1)$$

with $t_i = i/n$, $g \in L^2([0, 1])$ and ξ_i a Gaussian zero mean second order stationary process with long-range dependence, i.e. with autocovariances

$$\gamma(k) = E(\xi_i \xi_{i+k}) \underset{k \rightarrow \infty}{\sim} C_\gamma |k|^{-\alpha} \quad (2)$$

for some constants $\alpha \in (0, 1)$ and $C_\gamma > 0$ where “ $\sim$ ” means that the ratio of the two sides converges to one. For the spectral density $f(\lambda) = (2\pi)^{-1} \sum \gamma(k) \exp(-ik\lambda)$ this corresponds to a pole at the origin of the form $C_f |\lambda|^{\alpha-1}$ for a suitable constant C_f (see e.g. [6,78,68]). The issue addressed here is testing whether there is at least one jump in the trend function g . A natural approach is provided by wavelet analysis, since wavelets are likely to pick up even local deviations from smoothness. More specifically, as discussed in [13], a suitable wavelet estimator of $g(t)$ can be decomposed into a low resolution part $\hat{g}_{low}(t)$ characterizing the smooth component of g and a detailed high-resolution part $\hat{g}_{high}(t)$ that may capture jumps. This fact is used to develop a bootstrap based test for the null hypothesis that g is continuous everywhere against the alternative that g has at least one isolated jump.

In the iid and weakly dependent setting, the question of detecting changes or structural breaks in a mean or a (parametric or nonparametric) regression function has been considered by many authors. For an overview of the extended literature see

* Corresponding author.

E-mail address: jan.beran@uni-konstanz.de (J. Beran).

e.g. [38,19,5] and references therein. An interesting approach is taken in a recent paper by Ait-Sahalia and Jacod [2] who define an index of jump activity in a short-memory context (also see references therein for related literature). In the case of long memory, it has been recognized that structural breaks can be confused with stationary long-memory components, and standard tests do not apply [45,47,75,46,23,24,51,63,37,56,54,14,17,35]. In spite of this insight, only a few formal tests are available for the situation where the time point of the break is unknown ([75,51,54]; for an overview and more references see e.g. [69,5]). In particular, since sample paths of stationary processes with strong long memory tend to exhibit local spurious trends, very local changes in the mean function are difficult to detect. It is therefore very difficult to decide whether a completely unspecified (and thus nonparametric) trend function is continuous or only piecewise continuous. No general test for this hypothesis seems to be available in the long-memory setting. Due to their localization in time and frequency, wavelets provide a natural approach to solve this problem. This approach will be taken here.

Nonparametric regression with long-memory errors has been considered by several authors. Hall and Hart [40], Csörgö and Mielniczuk [20], Ray and Tsay [64], Robinson [65] and Beran and Feng [9,10] provide results for kernel regression. Other references are, for instance, [32–34] for nonparametric quantile smoothing, [11,12] for local polynomial estimation and [74,50,55,52,13] for wavelet thresholding. The literature on wavelet based trend estimation in the iid or weakly stationary case is much more abundant, some of the many important references being for instance [21,15,16,25–27,26,1,58,50,49,42–44,72,60,73,71]. While most results on wavelet thresholding under long memory focus on the minimax approach (e.g. [74,50]; also see [76] for random designs), Li and Xiao [55] and Beran and Shumeyko [13] derive data adaptive estimators, under the assumption that g or some derivatives of g are piecewise continuous. In the latter paper, it is shown that for functions with continuous derivatives, the rate given in Li and Xiao can be achieved without thresholding by choosing optimal values of J and q . This leads to a “low-resolution” estimator $\hat{g}_{low}$ for which exact constants for the MISE and asymptotic formulas for the optimal choice of the tuning parameters J and q (defined below) can be given explicitly [13]. This is comparable to optimal bandwidth selection in kernel smoothing [30,40,9,11]. To capture discontinuities, additional higher resolution levels combined with thresholding can be added so that the estimator can be written as $\hat{g} = \hat{g}_{low} + \hat{g}_{high}$.

In this paper, a bootstrap based test is developed for the null hypothesis that g is continuous everywhere against the alternative that g has at least one jump. In terms of the decomposition of $\hat{g}$ and g respectively, this amounts to testing the null hypothesis $H_0 : g_{high} \equiv 0$ against the alternative that $g_{high}(t)$ is not zero for at least one isolated point t . Properties of adaptive wavelet estimators of g_{low} and g_{high} are used to derive a test statistic W based on bootstrapped blocks of $Y_i - \hat{g}_{low}(t_i)$. Critical values of W under H_0 are obtained by suitable blockwise resampling of $Y_i - \hat{g}_{low}(t_i) - \hat{g}_{high}(t_i)$. Previous references to bootstrap in the long-memory setting include [53,41,39,59,22,29,4,54,3,70,62].

The paper is organized as follows. Basic definitions are introduced in Section 2, the main results are given in Section 3. The results are illustrated by a simulation study in Section 4. Final remarks in Section 5 conclude the paper. Proofs can be found in the Appendix.

2. Basic definitions

To include a large variety of wavelet generating functions (see [21,18]) we will use father and mother wavelets $\phi(t)$ and $\psi(t)$ respectively with compact support $[0, N]$ (for some $N \in \mathbb{N}$) instead of $[0, 1]$ only. The wavelet functions are assumed to be such that

$$\int_0^N \phi(t) dt = \int_0^N \phi^2(t) dt = \int_0^N \psi^2(t) dt = 1, \quad (3)$$

$$\psi(0) = \psi(N) = 0 \quad (4)$$

and, for any $J \geq 0$, the system $\{\phi_{jk}, \psi_{jk}, k \in \mathbb{Z}, j \geq 0\}$ with

$$\psi_{jk}(t) = N^{1/2} 2^{(j+j)/2} \psi(N 2^{j+j} t - k), \quad \phi_{jk}(t) = N^{1/2} 2^{j/2} \phi(N 2^j t - k),$$

is an orthonormal basis in $L^2(\mathbb{R})$. The number of vanishing moments of ψ will be denoted by $r \in \mathbb{N}$, i.e.

$$\int_0^N t^k \psi(t) dt = 0, \quad k = 0, 1, \dots, r-1 \quad (5)$$

and

$$\int_0^N t^r \psi(t) dt = v_r \neq 0. \quad (6)$$

Throughout the paper, ϕ and ψ will be assumed to be continuous and piecewise differentiable. Daubechies [21] (Chap. 6) provides examples of wavelets satisfying these conditions. For every $J \geq 0$, functions $g \in L^2[0, 1]$ have a unique orthogonal wavelet expansion

$$g(t) = \sum_{k=-N+1}^{N 2^J - 1} s_{jk} \phi_{jk}(t) + \sum_{j=0}^{\infty} \sum_{k=-N+1}^{N 2^{j+1} - 1} d_{jk} \psi_{jk}(t), \quad (7)$$

with wavelet coefficients

$$s_{jk} = \int_0^1 g(t) \phi_{jk}(t) dt, \quad d_{jk} = \int_0^1 g(t) \psi_{jk}(t) dt.$$

A (hard) thresholding wavelet estimator of g is defined by

$$\hat{g}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{jk} \phi_{jk}(t) + \sum_{j=0}^q \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk} I(|\hat{d}_{jk}| > \delta_j) \psi_{jk}(t) \quad (8)$$

where J , q and δ_j denote the decomposition level, smoothing parameter and threshold respectively, and the wavelet coefficients $\hat{s}_{jk}$ and $\hat{d}_{jk}$ are defined by

$$\hat{s}_{jk} = \frac{1}{n} \sum_{i=1}^n Y_i \phi_{jk}(t_i), \quad \hat{d}_{jk} = \frac{1}{n} \sum_{i=1}^n Y_i \psi_{jk}(t_i) \quad (9)$$

(see e.g. [25,21,1]). The estimator defined by Eq. (8) can be understood as a combination of two components, a smoothing component

$$\hat{g}_{low}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{jk} \phi_{jk}(t) \quad (10)$$

consisting of the scaling function decomposition only with no thresholding, and a higher resolution component

$$\hat{g}_{high}(t) = \sum_{j=0}^q \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk} I(|\hat{d}_{jk}| > \delta_j) \psi_{jk}(t) \quad (11)$$

where estimated coefficients are filtered by thresholding. The first component provides a good estimate of smooth functions whereas the second component is useful for modeling discontinuities (see [13]).

In the derivation of optimal data adaptive wavelet estimators, the following constants play an important role (see [13,55]),

$$C_\phi^2 = C_\gamma \int_0^N \int_0^N |x-y|^{-\alpha} \phi(x) \phi(y) dx dy, \quad (12)$$

$$C_\psi^2 = C_\gamma \int_0^N \int_0^N |x-y|^{-\alpha} \psi(x) \psi(y) dx dy. \quad (13)$$

The asymptotically optimal rate of the *MISE* of $\hat{g}$ defined in (8) is attained for

$$J = J_n = \frac{\alpha}{2r + \alpha} \log_2 n + C \quad \text{where } C \in \mathbb{R} \quad (14)$$

$$q = q_n = \log_2 n - J \quad (15)$$

and

$$2^{J+j} \delta_j^2 \rightarrow 0, \quad 2^{(J+j)(2r+1)} \delta_j^2 \rightarrow \infty, \quad \delta_j^2 \frac{n^\alpha 2^{(J+j)(1-\alpha)}}{\ln n} \rightarrow \infty \quad (16)$$

(see e.g. [55,13]). In the following we will use the notation $\kappa(n)$ for the sequence

$$\kappa(n) = \frac{1}{4} C_\psi^{-2} N^{\alpha-1} \delta_j^2 \frac{n^\alpha 2^{(J+j)(1-\alpha)}}{\ln n}. \quad (17)$$

Note that, by definition, $\kappa(n) \rightarrow \infty$.

Consider now the null hypothesis $H_0 : g \in C[0, 1]$ against the alternative that g is only piecewise continuous, with a finite number of isolated jumps. The following statistics will be used to define a bootstrap based test of H_0 . Denote by X_i the residuals after extracting the estimated trend $\hat{g}(t_i)$ given in (8),

$$X_i = Y_i - \hat{g}(t_i) = \xi_i + g(t_i) - \hat{g}_{low}(t_i) - \hat{g}_{high}(t_i).$$

For blocks of size l ($1 \leq l \leq n$) define block sums $\zeta_1, \dots, \zeta_{n-l+1}$ by

$$\zeta_i = X_i + \dots + X_{i+l-1} \quad (1 \leq i \leq n-l+1).$$

Then draw $\zeta_1^*, \dots, \zeta_k^*$ randomly with replacement from $\{\zeta_1, \dots, \zeta_{n-l+1}\}$ and calculate

$$T_{kl}^* = a_l^{-1} \left(k^{-1/2} \sum_{i=1}^k \zeta_i^* \right) \quad (18)$$

where $a_l = C_Y^{1/2} l^{1-\alpha/2}$. Similarly, residuals $X_{i,low}$ are defined by subtracting the low resolution component only, i.e.

$$X_{i,low} = Y_i - \hat{g}_{low}(t_i) = \xi_i + g(t_i) - \hat{g}_{low}(t_i)$$

and corresponding block sums are denoted by

$$\zeta_{i,low} = X_{i,low} + \dots + X_{i+l-1,low} \quad (1 \leq i \leq n-l+1).$$

Sampling randomly with replacement from $\{\zeta_{1,low}, \dots, \zeta_{n-l+1,low}\}$ we obtain $\zeta_{1,low}^*, \dots, \zeta_{k,low}^*$ and

$$T_{kl,low}^* = a_l^{-1} \left(k^{-1/2} \sum_{i=1}^k \zeta_{i,low}^* \right). \quad (19)$$

In the following section it is shown that under H_0 , T_{kl}^* and $T_{kl,low}^*$ have approximately the same distribution whereas this is no longer the case under H_1 .

3. Main results

3.1. Asymptotic distribution of T_{kl}^* and $T_{kl,low}^*$

Throughout the paper, E_n and Var_n will denote expectation and variance with respect to the distribution generated by the resampling procedure, conditionally on the observations $Y_1, \dots, Y_n$. Similarly, $P_n(A)$ will denote the conditional probability of A . For block lengths increasing like a power of n (but slower than n) and bootstrap sample size k tending to infinity, the expected value and variance of $T_{kl,low}^*$ can be approximated as follows. (Note that, since E_n , Var_n and P_n are random variables depending on $Y_1, \dots, Y_n$, errors of approximations are of the form $o_p(\cdot)$ instead of $o(\cdot)$.)

Theorem 1. Let $l = O(n^{1-\epsilon})$ with $1 - 2(2r + \alpha)^{-1} < \epsilon < 1 - (2r + \alpha)^{-1}$. Assume that $g \in L^2[0, 1]$, g' exists on $[0, 1]$ except for a finite number of points, and, where it exists, g' is piecewise continuous and bounded. Then the following hold as $\min\{k, n\} \rightarrow \infty$.

(a) If g has no jumps on $(0, 1)$, then

$$E_n(T_{kl,low}^*) = o_p(1)$$

and

$$Var_n(T_{kl,low}^*) = \tilde{\sigma}^2 + o_p(1)$$

where

$$\tilde{\sigma}^2 = 2\sigma^2(1 - \alpha)^{-1}(2 - \alpha)^{-1}. \quad (20)$$

(b) If g has at least one jump on $(0, 1)$, then

$$E_n(T_{kl,low}^*) = o_p(1)$$

and there exists a constant $C^* > 0$ such that

$$Var_n(T_{kl,low}^*) = \tilde{\sigma}^2 + v_n + o_p(1)$$

with

$$v_n = C^* n^{\alpha(1-\epsilon) - \frac{\alpha}{2r+\alpha}}.$$

Theorem 1 means that under H_1 , the variance of $T_{kl,low}^*$ increases by the additional term v_n . Since $\epsilon < 1 - (2r + \alpha)^{-1}$, v_n diverges to infinity as n increases. Note also that ϵ can be specified as required in **Theorem 1** without a priori knowledge of α , because $0 < \alpha < 1$ and $r \geq 1$ implies

$$1 - \frac{2}{2r + \alpha} < 1 - \frac{2}{2r + 1} < 1 - \frac{1}{2r} < 1 - \frac{1}{2r + \alpha}.$$

We may thus choose an ϵ from the interval $[1 - 2(2r + 1)^{-1}, 1 - 1/(2r)]$.

The following result shows that the standardized bootstrapped statistic $T_{kl,low}^*$ is asymptotically normal under H_0 and H_1 respectively and convergence is uniform on $\mathbb{R}$.

Theorem 2. Under the assumptions of **Theorem 1** the following hold:

(a) If g has no jumps on $(0, 1)$, then

$$\sup_{x \in \mathbb{R}} \left| P_n(T_{kl,low}^* \leq x) - \Phi\left(\frac{x}{\tilde{\sigma}}\right) \right| = o_p(1).$$

(b) If g has at least one jump on $(0, 1)$, then

$$\sup_{x \in \mathbb{R}} |P_n(v_n^{-1/2} T_{kl,low}^* \leq x) - \Phi(x)| = o_p(1)$$

where v_n is as in Theorem 1(b).

The next step is to compare the distributions of $T_{kl,low}^*$ and T_{kl}^* . First it is shown that under the null hypothesis both variables are asymptotically equivalent. However, in contrast to $T_{kl,low}^*$, the variance of T_{kl}^* is not affected by isolated jumps.

Theorem 3. Under the assumptions of Theorem 1 the following hold:

(a) If g has no jumps on $(0, 1)$, then

$$E_n(T_{kl}^*) - E_n(T_{low,kl}^*) = o_p(n^{\frac{1}{2}\alpha(1-\epsilon) - \frac{1}{4}\kappa(n)})$$

and

$$\text{Var}_n(T_{kl}^*) - \text{Var}_n(T_{low,kl}^*) = o_p(n^{\alpha(1-\epsilon) - \frac{1}{2}\kappa(n)}).$$

Moreover,

$$T_{kl}^* = T_{low,kl}^* + o_p(n^{\frac{1}{2}\alpha(1-\epsilon) - \frac{1}{4}\kappa(n)}).$$

(b) If g has at least one jump on $(0, 1)$, then

$$E_n(T_{kl}^*) = o_p(1)$$

and

$$\text{Var}_n(T_{kl}^*) = \tilde{\sigma}^2 + o_p(1).$$

Finally, it is shown that T_{kl}^* is also asymptotically normal, however, as implied by Theorem 3, with the same standardization under H_0 and H_1 .

Theorem 4. Under the assumptions of Theorem 1 the following hold.

(a) If g has no jumps on $(0, 1)$, then

$$\sup_{x \in \mathbb{R}} |P_n(T_{kl}^* \leq x) - \Phi(\tilde{\sigma}^{-1}x)| = o_p(1)$$

and

$$\sup_{x \in \mathbb{R}} |P_n(T_{kl}^* \leq x) - P_n(T_{kl,low}^* \leq x)| = o_p(1).$$

(b) If g has at least one jump on $(0, 1)$, then

$$\sup_{x \in \mathbb{R}} |P_n(T_{kl}^* \leq x) - \Phi(\tilde{\sigma}^{-1}x)| = o_p(1).$$

3.2. Testing for jumps

Given Theorems 1–4, a natural approach to testing for jumps is to compare $H_0 : \text{var}(T_{kl,low}^*) = \text{var}(T_{kl}^*)$ with $H_1 : \text{var}(T_{kl,low}^*) > \text{var}(T_{kl}^*)$. This can be done as follows. In a first step, m bootstrap samples $\{\zeta_{1,low}^{*(i)}, \dots, \zeta_{k,low}^{*(i)}\}$ ($i = 1, 2, \dots, m$) are drawn and $T_{kl,low}^{*(1)}, \dots, T_{kl,low}^{*(m)}$ are computed. The test statistic is defined by

$$W_{low} = \tilde{\sigma}^{-2} \sum_{i=1}^m (T_{kl,low}^{*(i)} - \bar{T}_{kl,low}^*)^2,$$

with $\bar{T}_{kl,low}^* = m^{-1} \sum_{i=1}^m T_{kl,low}^{*(i)}$. In a second step, critical values for W_{low} are calculated by resampling in the same way s sets of m bootstrapped values of T_{kl}^* , i.e. we have $A_1 = \{T_{kl}^{*(1,1)}, \dots, T_{kl}^{*(m,1)}\}, \dots, A_s = \{T_{kl}^{*(1,s)}, \dots, T_{kl}^{*(m,s)}\}$. For each A_j ($1 \leq j \leq s$), the statistic

$$W_j = \tilde{\sigma}^{-2} \sum_{i=1}^m (T_{kl}^{*(i,j)} - \bar{T}_{kl}^{*(j)})^2,$$

with $\bar{T}_{kl}^{*(j)} = m^{-1} \sum_{i=1}^m T_{kl}^{*(i,j)}$ is computed and critical values for W_{low} are obtained from the empirical distribution of $W_1, \dots, W_s$. The null hypothesis is rejected at the level of significance α , if W_{low} is above the $(1 - \alpha)$ -quantile. Theorems 1–4 imply that asymptotically, the level of the test is correct and the power converges to one.

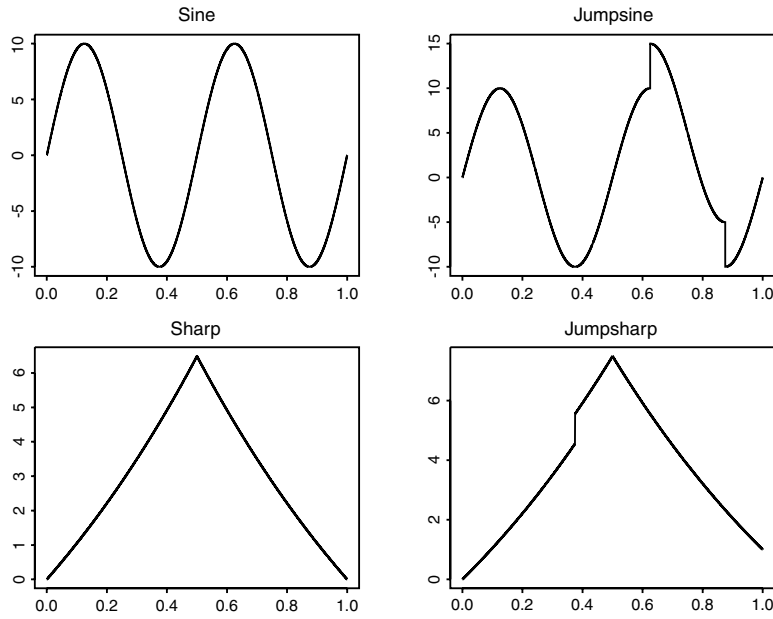


Fig. 1. Trend functions: (a) sine, (b) “JumpSine”, (c) “Sharp”, (d) “Sharp” with a single jump.

4. Simulations

The finite sample behavior of the proposed test is illustrated by the following simulations. In the main part of the simulation study, residuals are generated by a Gaussian FARIMA(0, d , 0) process (note that $\alpha = 1 - 2d$). Five trend functions g (Figs. 1 and 8(b)) are considered:

(a) Sine function: $g_1(t) = 10 \sin(4\pi t)$ (Fig. 1(a)).

(b) “JumpSine” function (Fig. 1(b)):

$$g_2(t) = 10 \sin(4\pi t) + \Delta \cdot I \left\{ \frac{5}{8} < t < \frac{7}{8} \right\}.$$

(c) “Sharp” function (Fig. 1(c)):

$$g_3(t) = 10[\exp(tI\{t < 0.5\} + (1-t)I\{t > 0.5\}) - 1].$$

(d) “JumpSharp” function, i.e. “Sharp” function with a single jump (Fig. 1(d)):

$$g_4(t) = 10[\exp(tI\{t < 0.5\} + (1-t)I\{t > 0.5\}) - 1] + \Delta I\{t > 3/8\}, \quad (\Delta > 0).$$

(e) Sharp plus a local constant shift (Fig. 8(b)):

$$g_5(t) = 10[\exp(tI\{t < 0.5\} + (1-t)I\{t > 0.5\}) - 1] + \Delta I \left\{ \frac{20}{64} \leq t \leq \frac{21}{64} \right\}, \quad (\Delta > 0).$$

Moreover, modifications of function (b) by more than one jump are also included (see below). The parameters q , J and δ_j were chosen as outlined in [13].

Simulated sample paths of Y_i with trend functions (a) and (b) are shown in Fig. 2(a) and (b). For the series in Fig. 2(b), the complete estimate $\hat{g}$ and the low resolution estimate $\hat{g}_{low}$ are displayed in Fig. 3(c) and (d) respectively. In this case, including the high resolution component mainly leads to the detection of the second jump. Fig. 4(a) and (b) show the residuals $X_{i,low} = Y_i - \hat{g}_{low}$ obtained after subtracting the low resolution estimate. The corresponding block sums $\zeta_{i,low}$ are shown in Fig. 5, and bootstrapped samples of $T_{kl,low}^{*(j)}$ (standardized by $\tilde{\sigma}$) are displayed in Fig. 6(a) and (b) respectively. Compared to Fig. 6(a), the variability of $T_{kl,low}^{*(j)}$ appears to be higher in Fig. 6(b). This is due to the discontinuity of the second trend function, as explained by Theorems 1(b) and 4(b).

Numerical results are summarized in Tables 2–14. An overview of the simulated models and parameter settings is given in Table 1. Tables 2 and 3 show simulated rejection probabilities for the sine and JumpSine function and $d = 0.1$ and 0.4 respectively. Results for different values of k are compared. The nominal levels of significance are achieved with a reasonable degree of accuracy for all cases, even for the relatively small sample size of $n = 128$ and strong long memory ($d = 0.4$). Similar results are obtained for significance levels for case (c) (Tables 8 and 9). As expected, the power increases with increasing values of k and n . Moreover, discontinuities are more difficult to detect under stronger long memory. This is not

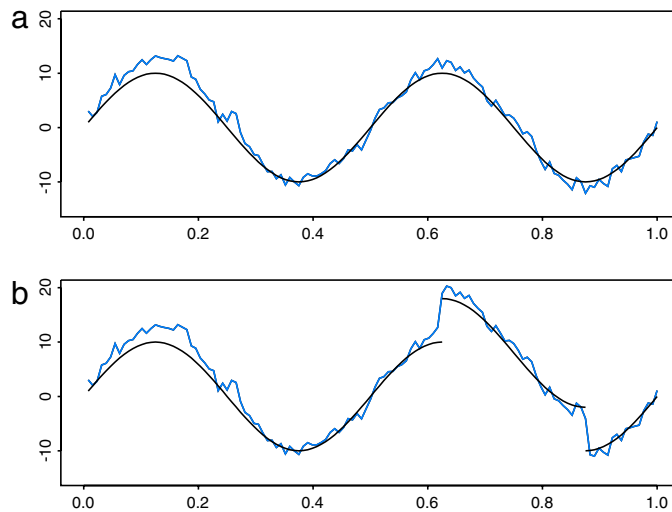


Fig. 2. Simulated data of length $n = 128$ with the trend functions (a) and (b) (sine and “JumpSine”) respectively and a FARIMA(0, 0.4, 0) residual process. In (b), Δ is equal to 8.

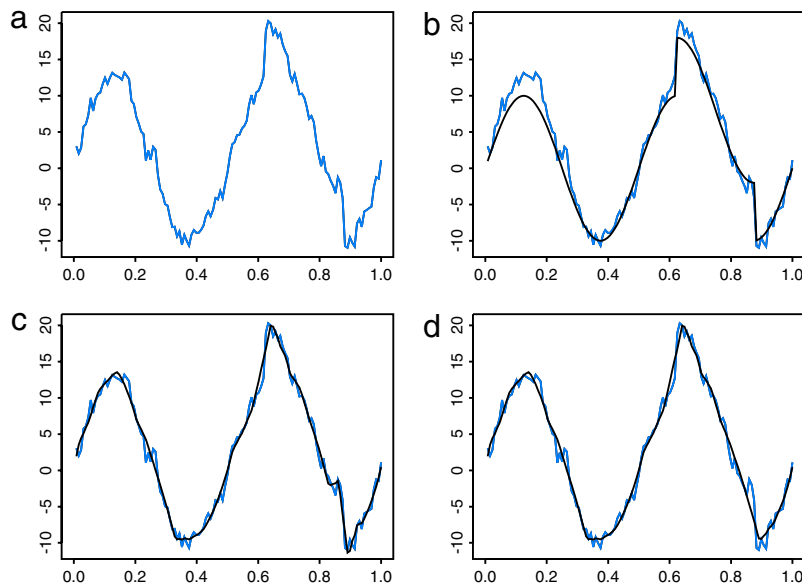


Fig. 3. Comparison of trend estimation by the low resolution component $\hat{g}_{low}$ and by a full wavelet decomposition $\hat{g} = \hat{g}_{low} + \hat{g}_{high}$ for the simulated series in Fig. 2(a) and (b) respectively. The wavelet basis “s6” was used. The figures are (a) data, (b) data and trend function, (c) data and $\hat{g}$, (d) data and $\hat{g}_{low}$.

surprising, since a large value of d leads to sample paths that tend to have local cycles and trends resembling discontinuities in the expected value.

The influence of Δ is illustrated in Table 4. In Table 5, it can be seen that the power increases considerably when the number of jumps is increased. Moreover, the power is also influenced by the choice of the wavelet basis. In the case of the JumpSine function, the power improves when one replaces “s6” by “s8” or “s10” respectively (Table 6). The effect of choosing different values of l (or ϵ) is considered in Table 7. As a general rule l should be neither too small nor too large. Choosing l very small (ϵ large) prevents the procedure from capturing the asymptotic effect of long memory. On the other hand, if l is very large (ϵ small), then local jumps are smoothed out and the number of available blocks is small.

Simulated powers for case (d) (sharp function with one jump) with $\Delta = 2$ are displayed in Tables 8 and 9. Case (e) is considered in Table 10. Discontinuities of this type are very difficult to detect, since under long memory local shifts are often masked by similar spurious local trends generated by the residual process. This is illustrated in Fig. 7. Looking at the data only, without any visual guideline, it is very difficult to see whether and where a jump in the mean function may have occurred. Fig. 8(c) illustrates that the high resolution fit detects the local jump, though the estimate is more wiggly than the true trend. The reason is that for $n = 512$ and relatively strong long memory with $d = 0.3$, the extremely local shift is

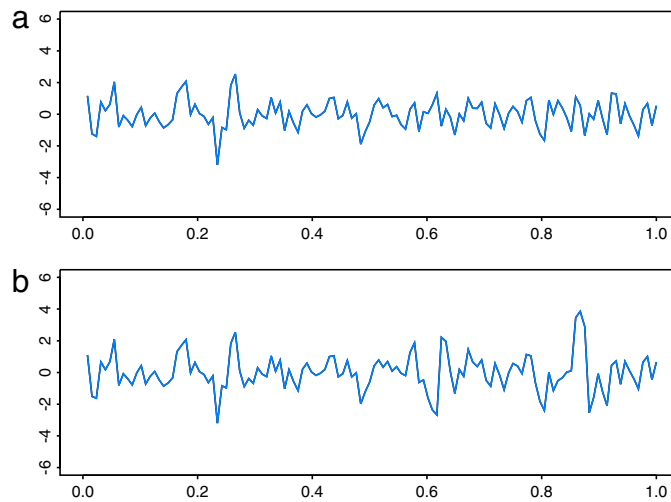


Fig. 4. Residuals for simulated data in Fig. 2(a) and (b) respectively after subtracting the low resolution estimate $\hat{g}_{low}$.

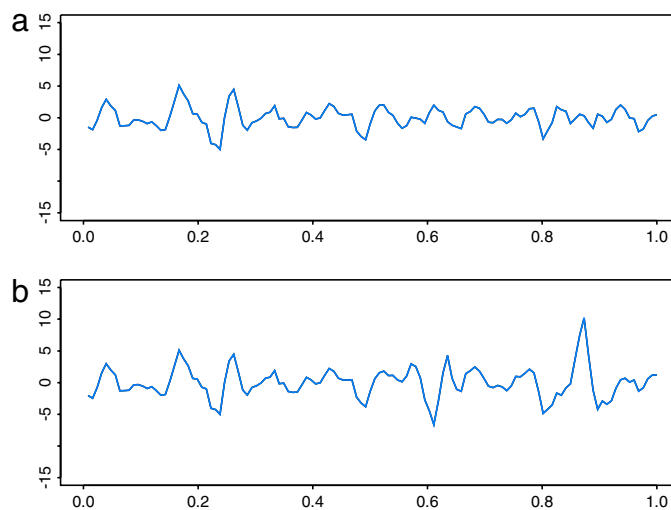


Fig. 5. Block sums $\zeta_{i,low}$ for the simulated data in Fig. 2(a) and (b) respectively.

Table 1

Overview of tables with simulation results.

Table	FARIMA model	ϵ	k	Δ	No. of jumps	Wavelet
Sine, JumpSine						
2 & 3	$d = 0.1, 0.4$	$1.5(2r + \alpha)^{-1}$	10–150	0, 10	2	s6
4	$d = 0.4$	$1.5(2r + \alpha)^{-1}$	50	5–12	2	s6
5	$d = 0.4$	$1.5(2r + \alpha)^{-1}$	50	5	1–4	s6
6	$d = 0.4$	$1.5(2r + \alpha)^{-1}$	50	10	2	s6, s8, s10
7	$d = 0.4$	$1 - 2(2r + 1)^{-1}, 1 - 1.5(2r + \alpha)^{-1}, 1 - \frac{1}{2}r^{-1}$	50	10	2	s6
Sharp, JumpSharp						
8 & 9	$d = 0.1, 0.4$	$1.5(2r + \alpha)^{-1}$	10–150	0, 2	1	s6
11 & 12	$d = 0.1, 0.4, \phi = -0.5$	$1.5(2r + \alpha)^{-1}$	10–150	0, 2	1	s6
13 & 14	$d = 0.1, 0.4,$ $\phi = 0.5, \psi = 0.5$	$1.5(2r + \alpha)^{-1}$	10–150	0, 2	1	s6
Sharp with local shift						
10	$d = 0.3$	$1.5(2r + \alpha)^{-1}$	150	3	2	s6

masked by spurious trends of the residual process. Nevertheless, it is remarkable that the location is identified correctly, in spite of several similar looking (but spurious) “discontinuities” at other time points. The simulation results in Table 10 show

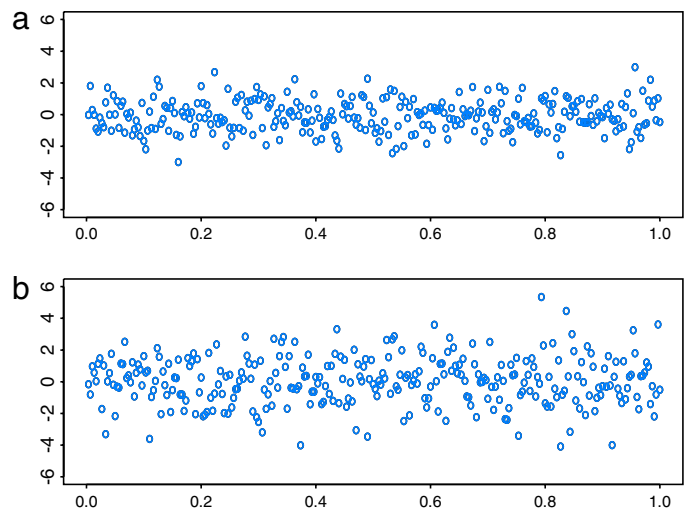


Fig. 6. Random sample $\{T_{kl,low}^{*(1)}, \dots, T_{kl,low}^{*(m)}\}$ for the simulated series in Fig. 2(a) and (b) respectively, standardized by $\bar{\sigma}$.

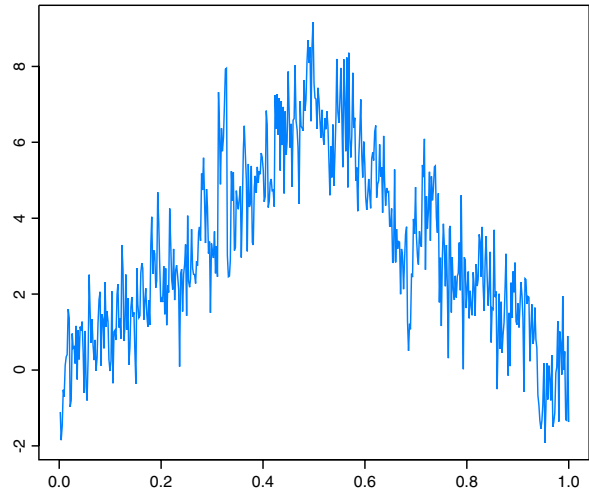


Fig. 7. Simulated data with trend function (e) (“Sharp” plus a local shift) with $(\Delta = 3)$ and residuals generated by a $FARIMA(0, 0.3, 0)$ process with sample size $n = 512$.

Table 2
Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.1, 0)$ process with g equal to the sine function (a) (upper part of the table) and the JumpSine function (b) with $\Delta = 10$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	$n = 128$		$n = 256$		$n = 512$	
	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.1075	0.0425	0.1050	0.0475	0.1050	0.0575
20	0.1075	0.0350	0.0925	0.0550	0.0950	0.0550
50	0.1000	0.0475	0.0925	0.0575	0.1050	0.0525
100	0.1050	0.0450	0.0950	0.0525	0.1050	0.0500
150	0.0950	0.0500	0.0975	0.0525	0.1025	0.0525
$\Delta = 10$						
10	0.8200	0.7575	0.9625	0.9250	0.9875	0.9650
20	0.9625	0.9050	0.9900	0.9850	1	1
50	1	0.9950	1	1	1	1
100	1	1	1	1	1	1
150	1	1	1	1	1	1

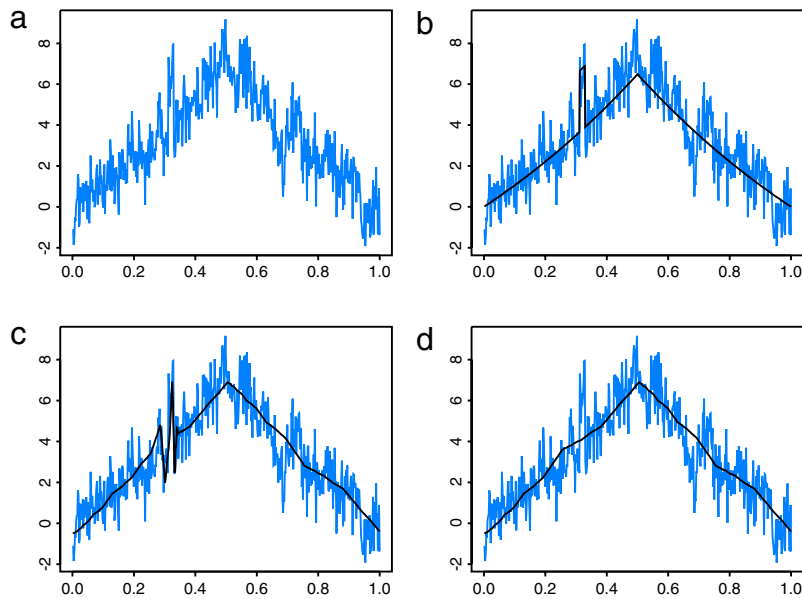


Fig. 8. Comparison of trend estimation by $\hat{g}_{low}$ and the full wavelet decomposition $\hat{g} = \hat{g}_{low} + \hat{g}_{high}$ for the simulated data in Fig. 7. The “s6” wavelet basis was used. The figures are (a) data, (b) data with trend function, (c) data with $\hat{g}$ and (d) data with $\hat{g}_{low}$.

Table 3

Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process with g equal to the sine function (a) (upper part of the table) and the JumpSine function (b) with $\Delta = 10$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.1275	0.0625	0.0950	0.0575	0.1075	0.0625
20	0.1000	0.0500	0.0975	0.0550	0.0950	0.0500
50	0.0950	0.0525	0.1025	0.0525	0.1050	0.0425
100	0.1000	0.0625	0.1050	0.0450	0.0900	0.0550
150	0.0975	0.0525	0.1025	0.0525	0.1000	0.0500
$\Delta = 10$						
10	0.3650	0.2975	0.4650	0.3775	0.6075	0.4950
20	0.5050	0.4075	0.6550	0.5600	0.6775	0.6250
50	0.7425	0.6925	0.7575	0.7125	0.7975	0.7575
100	0.8225	0.7750	0.8425	0.8050	0.8725	0.8225
150	0.8850	0.8375	0.9025	0.8750	0.9100	0.8675

Table 4

Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process of length $n = 128$ and g equal to the JumpSine function (b). The wavelet basis “s6” was used and k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400. Jumps of sizes $\Delta \in \{5, 6, \dots, 12\}$ are considered.

Δ	$\alpha = 0.1$	$\alpha = 0.05$
5	0.1100	0.0625
6	0.1525	0.0725
7	0.1725	0.0900
8	0.3075	0.2725
9	0.5825	0.4950
10	0.7425	0.6925
11	0.9550	0.9325
12	1	1

a reasonably good power of the bootstrap test already for $n = 512$. However, the power appears to converge to 1 rather slowly.

Finally, Tables 11–14 consider the same situation as in Tables 8 and 9, except that short-memory parameters are included. In Tables 11 and 12, we have a negative autoregressive parameter $\phi_1 = -0.5$ whereas strong positive short-

Table 5

Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process of length $n = 128$, and g equal to a JumpSine function with one or more jumps of size $\Delta = 5$. The wavelet basis “s6” was used and k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400. The following locations of the jumps were used: (1) one jump at 7/8; (2) two jumps at 5/8 and 7/8; (3) three jumps at 3/8, 5/8 and 7/8; and (4) four jumps at 1/8, 3/8, 5/8 and 7/8.

# Jumps	$\alpha = 0.1$	$\alpha = 0.05$
1	0.1525	0.1050
2	0.7250	0.6750
3	0.9800	0.9650
4	1	1

Table 6

Comparison of the wavelet basis functions “s6”, “s8” and “s10”. Given are empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process of length $n = 128$, and g equal to a JumpSine function with $\Delta = 10$. The value of k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

Basis	$\alpha = 0.1$	$\alpha = 0.05$
“s6”	0.7425	0.6925
“s8”	0.9100	0.8950
“s10”	0.9200	0.9025

Table 7

Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process for different values of ϵ ($l = n^{1-\epsilon}$), and g equal to the JumpSine function (b) with $\Delta = 10$. The wavelet basis “s6” was used and k was set equal to 50. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

n	$\epsilon = 1 - \frac{2}{2r+1}$		$\epsilon = 1 - \frac{1.5}{2r+\alpha}$		$\epsilon = 1 - \frac{1}{2r}$	
	0.1	0.05	0.1	0.05	0.1	0.05
2^7	0.6450	0.5825	0.7275	0.7075	0.7025	0.6825
2^8	0.7725	0.7000	0.7450	0.7125	0.7525	0.7050
2^9	0.7750	0.7000	0.8050	0.7550	0.7775	0.7100

Table 8

Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.1, 0)$ process with g equal to the Sharp function (c) (upper part of the table) and the JumpSharp function (d) with $\Delta = 2$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	$n = 128$		$n = 256$		$n = 512$	
	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.0900	0.0400	0.0950	0.0575	0.1075	0.0425
20	0.1050	0.0575	0.0950	0.0500	0.0950	0.0475
50	0.1050	0.0500	0.0975	0.0525	0.1075	0.0425
100	0.1025	0.0575	0.0950	0.0475	0.0975	0.0575
150	0.0975	0.0425	0.1075	0.0450	0.1025	0.0550
$\Delta = 2$						
10	0.4750	0.3650	0.6250	0.5475	0.7225	0.6375
20	0.6175	0.5225	0.8425	0.7325	0.8925	0.8450
50	0.8150	0.7150	0.9600	0.9450	0.9825	0.9600
100	0.9225	0.8750	0.9825	0.9825	1	1
150	0.9575	0.9525	1	1	1	1

range dependence is added in Tables 13 and 14 with $\phi_1 = 0.5$ and $\psi = 0.5$ (Table 14). The results indicate that a negative short-memory component improves power. This coincides with the fact that negative autocorrelations improve the accuracy of trend estimates. For the $FARIMA(1, d, 1)$ model with positive short-memory components the results are less clear. Here, the choice of k apparently plays a role.

5. Final remarks

The test proposed here is fully nonparametric in the sense that the tuning constants ϵ and k can be chosen without knowledge of the underlying process. The only unknown parameters that would need to be estimated from the observed

Table 9

Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.4, 0)$ process with g equal to the Sharp function (c) (upper part of the table) and the JumpSharp function (d) with $\Delta = 2$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.0950	0.0550	0.1000	0.0475	0.0950	0.0425
20	0.0950	0.0575	0.0950	0.0425	0.0975	0.0425
50	0.0975	0.0550	0.0975	0.0425	0.1050	0.0525
100	0.0950	0.0550	0.1025	0.0425	0.1025	0.0450
150	0.1000	0.0550	0.0950	0.0500	0.0975	0.0550
$\Delta = 2$						
10	0.1350	0.1000	0.2250	0.1550	0.2650	0.1900
20	0.2100	0.1400	0.3225	0.2450	0.3125	0.2275
50	0.2550	0.2050	0.3075	0.2550	0.4450	0.3700
100	0.2700	0.2300	0.3300	0.2850	0.4325	0.3925
150	0.2775	0.2450	0.3475	0.3125	0.5200	0.4850

Table 10

Empirical rejection frequencies based on 400 simulations of a $FARIMA(0, 0.3, 0)$ residual process and g equal to the Sharp function with a local shift (of size $\Delta = 3$ and length $1/64$) as defined in (e). The value of k was set equal to 150, the wavelet basis was “s6”. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

n	$\alpha = 0.1$	$\alpha = 0.05$
2^9	0.3400	0.2850
2^{10}	0.4050	0.3150
2^{12}	0.4175	0.3225
2^{15}	0.6450	0.4900

Table 11

Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.1, 0)$ process with $\phi = \{-0.5\}$, and g equal to the Sharp function (c) (upper part of the table) and the JumpSharp function (d) with $\Delta = 2$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05. For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	$n = 128$		$n = 256$		$n = 512$	
	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.0975	0.0625	0.0900	0.0725	0.0750	0.0500
20	0.1250	0.0500	0.0700	0.0375	0.0850	0.0550
50	0.1275	0.0800	0.1250	0.0400	0.0975	0.0475
100	0.1100	0.0400	0.1175	0.0450	0.0800	0.0625
150	0.1125	0.0625	0.0925	0.0550	0.0925	0.0500
$\Delta = 2$						
10	0.4150	0.3275	0.7100	0.6425	0.7800	0.6925
20	0.3425	0.2850	0.8650	0.7300	0.9150	0.8800
50	0.6025	0.5000	0.9775	0.9600	1	1
100	0.5975	0.5750	1	1	1	1
150	0.6550	0.6475	1	1	1	1

data are C_γ and α . In principle, these parameters can be estimated by applying parametric or semiparametric methods to the estimated residuals $\hat{X}_i$ (see e.g. [28,36,7,8,31,66,67,48,57]). However, the optimal choice of the tuning parameters J and q for estimating g depends on α and C_γ . In analogy to [9,10] or [64], a data adaptive iterative algorithm for estimating g , α and C_γ simultaneously can be developed.

Acknowledgments

This work was supported in part by a grant of the German Research Foundation (Research Unit 518). We would like to thank the associate editor and two referees for their constructive comments that lead to an improved presentation of the results. We would also like to thank Don Percival (University of Washington) for interesting discussions that inspired us to look at bootstrap in the context of wavelets.

Table 12

Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.4, 0)$ process with $\phi = \{-0.5\}$, and g equal to the Sharp function (c) (upper part of the table) and the JumpSharp function (d) with $\Delta = 2$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05 . For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	$n = 128$		$n = 256$		$n = 512$	
	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.0975	0.0625	0.0900	0.0725	0.0750	0.0500
20	0.1250	0.0500	0.0700	0.0375	0.0850	0.0550
50	0.1275	0.0800	0.1250	0.0400	0.0975	0.0475
100	0.1100	0.0400	0.1175	0.0450	0.0800	0.0625
150	0.1125	0.0625	0.0925	0.0550	0.0925	0.0500
$\Delta = 2$						
10	0.4650	0.4125	0.5200	0.4575	0.6275	0.5400
20	0.5600	0.5125	0.6000	0.5325	0.6625	0.6325
50	0.7425	0.6600	0.7625	0.7225	0.7700	0.7225
100	0.7725	0.7550	0.8025	0.7700	0.8725	0.8450
150	0.8025	0.7700	0.8725	0.8575	0.8900	0.8675

Table 13

Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.1, 1)$ process with $\phi = \{0.5\}$ and $\psi = \{0.5\}$, and g equal to the Sharp function (c) (upper part of the table) and the JumpSharp function (d) with $\Delta = 2$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05 . For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	$n = 128$		$n = 256$		$n = 512$	
	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.1175	0.0600	0.1475	0.0600	0.0800	0.0325
20	0.1150	0.0525	0.1125	0.0625	0.1150	0.0600
50	0.0800	0.0400	0.1525	0.0475	0.1175	0.0425
100	0.1225	0.0775	0.1050	0.0575	0.1075	0.0475
150	0.1150	0.0550	0.1150	0.0550	0.1000	0.0525
$\Delta = 2$						
10	0.4300	0.3475	0.6375	0.5475	0.7175	0.6325
20	0.6150	0.5600	0.8475	0.7600	0.9125	0.8850
50	0.8225	0.7650	0.9975	0.9850	1	1
100	0.8975	0.8950	1	1	1	1
150	0.9225	0.9100	1	1	1	1

Table 14

Empirical rejection frequencies based on 400 simulations of a $FARIMA(1, 0.4, 1)$ process with $\phi = \{0.5\}$ and $\psi = \{0.5\}$, with g equal to the Sharp function (c) (upper part of the table) and the JumpSharp function (d) with $\Delta = 2$ (lower part of the table). The wavelet basis “s6” was used. The levels of significance are $\alpha = 0.1$ and 0.05 . For each simulation, the rejection regions were determined using a bootstrap sample of size 400.

$k \setminus \alpha$	$n = 128$		$n = 256$		$n = 512$	
	0.1	0.05	0.1	0.05	0.1	0.05
$\Delta = 0$						
10	0.1250	0.0800	0.0900	0.0750	0.1250	0.0500
20	0.1475	0.0875	0.1325	0.0850	0.1000	0.0275
50	0.1375	0.0650	0.0950	0.0325	0.1425	0.0425
100	0.1100	0.0875	0.1075	0.0600	0.1475	0.0800
150	0.0850	0.0675	0.1000	0.0525	0.1075	0.0625
$\Delta = 2$						
10	0.2475	0.1650	0.2300	0.1400	0.2425	0.1575
20	0.2800	0.2475	0.2325	0.1875	0.2650	0.2125
50	0.3325	0.2775	0.3625	0.2850	0.3800	0.3650
100	0.3650	0.3200	0.4025	0.3775	0.5450	0.5050
150	0.4650	0.4400	0.4150	0.3875	0.5900	0.5200

Appendix

A.1. Proofs

The following assumptions will be used.

(A1) The derivatives ψ' , ϕ' and $g^{(r)}$ exist except for a finite number of points and, where they exist, they are piecewise continuous and bounded.

(A2) $J, j \geq 0$ and $-(N-1) \leq k \leq N2^{J+j} - 1$.

(A3)

$$\begin{aligned} v_k &= \int_0^N t^k \psi(t) dt = 0 \quad (0 \leq k \leq r-1), \\ v_r &\neq 0. \end{aligned}$$

In the case where g is constant and estimated by the sample mean, the results under H_0 essentially follow from [53]. In the following these results are extended to wavelet estimation of a nonconstant function g , under H_0 and H_1 respectively. At first, we cite results that follow from [13,55].

Lemma 1. Under (A1)–(A3) the following hold:

- (a) $s_{jk} = O(2^{-J/2})$;
- (b) $E(\hat{s}_{jk}) - s_{jk} = O(n^{-1}2^{J/2})$;
- (c) $E\{[\hat{s}_{jk} - E(\hat{s}_{jk})]^2\} = C_\phi^2 N^{-1+\alpha} n^{-\alpha} 2^{-J(1-\alpha)} + O(n^{-1})$;
- (d) $\sigma_j^2 = E[(\hat{d}_{jk} - E(\hat{d}_{jk}))^2] = C_\psi^2 N^{-(1-\alpha)} n^{-\alpha} 2^{-(J+j)(1-\alpha)} + O(n^{-1})$;
- (e) $E[\hat{d}_{jk}] - d_{jk} = O(n^{-1}2^{(J+j)/2})$. (21)

(f) If $g^{(r)} \in C(I_{jk})$ where $I_{jk} = [kN^{-1}2^{-(J+j)}, (k+N)N^{-1}2^{-(J+j)}]$, then

$$d_{jk} = \frac{v_r}{r!} g^{(r)}(kN^{-1}2^{-(J+j)}) N^{-\frac{2r+1}{2}} 2^{-\frac{2r+1}{2}(J+j)} + o(2^{-\frac{2r+1}{2}(J+j)}). \quad (22)$$

(g) If there exists an $m \in \{0, 1, \dots, r-1\}$ such that $g \in C^m(I_{jk})$, $g^{(m+1)}$ exists almost everywhere, but $g \notin C^{m+1}(I_{jk})$, then

$$d_{jk} = O(2^{-\frac{2(m+1)+1}{2}(J+j)}).$$

(h) If $g \notin C(I_{jk})$, then

$$d_{jk} = O(2^{-(J+j)/2}).$$

Proof. (a) through (f) follow directly from [13,55].

For (g), we have

$$\begin{aligned} d_{jk} &= N^{1/2} 2^{(J+j)/2} \int_0^1 g(t) \psi(N2^{J+j}t - k) dt \\ &= N^{-1/2} 2^{-(J+j)/2} \int_0^N g(N^{-1}2^{-(J+j)}(y+k)) \psi(y) dy. \end{aligned} \quad (23)$$

The points where g is not differentiable are isolated so that, for n large enough, there is exactly one such point in the support of ψ_{jk} . Thus, assume that there exists exactly one y^* such that $y_{jk}^* = N^{-1}2^{-(J+j)}(y^* + k) \in \text{supp}(\psi_{jk})$ and

$$\lim_{\Delta \rightarrow 0} [g^{(m+1)}(y_{jk}^* + N^{-1}2^{-(J+j)}\Delta) - g^{(m+1)}(y_{jk}^* - N^{-1}2^{-(J+j)}\Delta)] = C_{m+1, y^*} \neq 0.$$

Since g is m -times continuously differentiable on I_{jk} , and $m + 1$ -times continuously differentiable on $I_{jk} \setminus \{y_{jk}\}$, a local Taylor expansion of g at y^* (see e.g. [77] pp. 225–226), yields

$$d_{jk} = (N2^{(j+j)})^{-\frac{1}{2}} \sum_{u=0}^m \frac{(N2^{(j+j)})^{-u}}{u!} g^{(u)}(y_{jk}^*) \int_0^N \psi(y)(y - y^*)^u dy + \frac{(N2^{(j+j)})^{-(m+3/2)}}{(m+1)!} (A^- + A^+) + o(2^{-(m+3/2)(j+j)})$$

with

$$A^- = g^{(m+1)}(y_{jk}^* - 0) \int_0^{y^*} \psi(y)(y - y^*)^{m+1} dy,$$

$$A^+ = g^{(m+1)}(y_{jk}^* + 0) \int_{y^*}^N \psi(y)(y - y^*)^{m+1} dy.$$

Taking into account condition (5) and $m < r$, we have

$$\begin{aligned} A^- + A^+ &= g^{(m+1)}(y_{jk}^* - 0) \int_0^N \psi(y)(y - y^*)^{m+1} dy + [g^{(m+1)}(y_{jk}^* + 0) - g^{(m+1)}(y_{jk}^* - 0)] \int_{y^*}^N \psi(y)(y - y^*)^{m+1} dy \\ &= g^{(m+1)}(y_{jk}^* - 0) v_{m+1} + C_{m+1, y^*} \int_{y^*}^N \psi(y)(y - y^*)^{m+1} dy \end{aligned}$$

and hence

$$d_{jk} = 2^{-\frac{2(m+1)+1}{2}(j+j)} C_{jk} + o(2^{-(m+3/2)(j+j)}) = O(2^{-\frac{2(m+1)+1}{2}(j+j)})$$

where C_{jk} is a finite constant.

Finally note that (h) follows directly from (23). $\square$

Now we consider the expected value and the autocovariance function of $\hat{g}_{low}$ under the null hypothesis.

Lemma 2. Suppose that $g \in C[0, 1]$, and (A1)–(A3) hold. Then

$$Z_i = \hat{g}_{low}(t_i) - E[\hat{g}_{low}(t_i)] \sim N(0, \sigma_i^2)$$

and

$$E[\hat{g}_{low}(t_i)] = g(t_i) + O(n^{-\frac{\alpha}{2r+\alpha}})$$

where

$$\sigma_i^2 \leq C_1 n^{-\frac{2r\alpha}{2r+\alpha}}$$

and, as $|k| \rightarrow \infty$,

$$\text{cov}(Z_i, Z_{i+k}) \leq C_2 |k|^{-\alpha}$$

for some positive constants C_1 and C_2 .

Proof. Taking into account (7), (9), (10), Lemma 1(b) and the bounded support of ϕ_{jk} we obtain

$$\begin{aligned} g(t_i) - \hat{g}_{low}(t_i) &= g(t_i) - E[\hat{g}_{low}(t_i)] + E[\hat{g}_{low}(t_i)] - \hat{g}_{low}(t_i) \\ &= \sum_{k=-(N-1)}^{N2^j-1} (s_{jk} - E[\hat{s}_{jk}]) \phi_{jk}(t_i) + \sum_{j=0}^{\infty} \sum_{k=-(N-1)}^{N2^{j+j}-1} d_{jk} \psi_{jk}(t_i) + \sum_{k=-(N-1)}^{N2^j-1} (E[\hat{s}_{jk}] - \hat{s}_{jk}) \phi_{jk}(t_i) \\ &= O(2^j n^{-1}) + \sum_{j=0}^{\infty} \sum_{k=-(N-1)}^{N2^{j+j}-1} d_{jk} \psi_{jk}(t_i) - n^{-1} \sum_{u=1}^n \xi_u \sum_{k=-(N-1)}^{N2^j-1} \phi_{jk}(t_u) \phi_{jk}(t_i). \end{aligned}$$

The stochastic part is normal and can be written as

$$\begin{aligned} Z_i &= n^{-1} \sum_{u=i-n2^{-j}}^{i+n2^{-j}} \xi_u \sum_{k=N2^j n^{-1} i-N}^{N2^j n^{-1} i} \phi_{jk}(t_u) \phi_{jk}(t_i) \\ &= N2^j n^{-1} \sum_{u=i-n2^{-j}}^{i+n2^{-j}} \xi_u \sum_{k=N2^j n^{-1} i-N}^{N2^j n^{-1} i} \phi(N2^j u n^{-1} - k) \phi(N2^j i n^{-1} - k) \\ &= N2^j n^{-1} \sum_{u=i-n2^{-j}}^{i+n2^{-j}} b_{ui} \xi_u \end{aligned}$$

where b_{ui} are some real constants. By analogous arguments as in [13,55] one obtains

$$\begin{aligned} \text{Var}(Z_i) &= N^2 n^{-2} 2^{2J} \text{Var} \left(\sum_{u=i-n2^{-J}}^{i+n2^{-J}} b_{ui} \xi_u \right) \\ &\leq N^2 n^{-2} 2^{2J} \max_{u,i} \{b_{ui}^2\} \left[\sum_{u=i-n2^{-J}}^{i+n2^{-J}} \text{Var}(\xi_u) + \sum_{\substack{u,v=i-n2^{-J} \\ u \neq v}}^{i+n2^{-J}} |\gamma(u-v)| \right] \\ &\sim 2N^2 \max_{u,i} \{b_{ui}^2\} (n2^{-J})^{-\alpha} C_\gamma \int_0^1 \int_0^1 |x-y|^{-\alpha} dx dy + O(n^{-1} 2^J). \end{aligned}$$

Eq. (14) then implies, for n large enough,

$$\text{Var}(Z_i) \leq C_1 n^{-\frac{2r\alpha}{2r+\alpha}} \quad (24)$$

for all i , where C_1 is a suitable constant. Finally Lemma 1(f) implies

$$\begin{aligned} \left| \sum_{j=0}^{\infty} \sum_{k=0}^{N2^{J+j}-1} d_{jk} \psi_{jk}(t_i) \right| &\leq \sum_{j=0}^{\infty} \sum_{k=N2^{J+j}n^{-1}i-N}^{N2^{J+j}n^{-1}i} \max_k |d_{jk}| \max_k |\psi_{jk}(t_i)| \\ &\leq C_3 \sum_{j=0}^{\infty} 2^{-\frac{3}{2}(J+j)} 2^{\frac{1}{2}(J+j)} = O(2^{-J}) = O(n^{-\frac{\alpha}{2r+\alpha}}) \end{aligned}$$

where $C_3 > 0$ is a suitable constant. In summary,

$$E[g(t_i) - \hat{g}_{low}(t_i)] = O(n^{-\frac{\alpha}{2r+\alpha}}),$$

$$\text{Var}[\hat{g}_{low}(t_s)] = O(n^{-\frac{2r\alpha}{2r+\alpha}}).$$

Finally,

$$\begin{aligned} \text{cov}(Z_i, Z_j) &\leq \max_{u,i} \{b_{u,i}^2\} N^2 n^{-2} 2^{2J} \sum_{u=i-n2^{-J}}^{i+n2^{-J}} \sum_{v=j-n2^{-J}}^{j+n2^{-J}} |E(\xi_u \xi_v)| \\ &\sim C_4 n^{-2} 2^{2J} \cdot (n2^{-J})^{2-\alpha} \int_{-1}^1 \int_{-1}^1 \left| \frac{i-j}{n2^{-J}} + x - y \right|^{-\alpha} dx dy + O(n^{-(1+\alpha)} 2^{(1+\alpha)J} |i-j|^{-(1+\alpha)}) \\ &\leq C_5 |i-j|^{-\alpha} \end{aligned}$$

for suitable constants C_4, C_5 and $|i-j| \rightarrow \infty$. $\square$

The next Lemma provides a useful approximation of $\hat{g}_{high}$.

Lemma 3. Suppose that (A1)–(A3) hold. Define for each j the set $D_j = \{d_{jk} : g \in C[\text{supp}(\psi_{jk})]\}$. Then

$$\hat{g}_{high}(t_i) = \sum_{j=0}^q \sum_{k=-(N-1)}^{N2^{J+j}-1} \hat{d}_{jk} I\{d_{jk} \in \bar{D}_j\} \psi_{jk}(t_s) + \eta(t_i)$$

where for any fixed $M \in \mathbb{N}$ and $\omega \in (0, 1)$,

$$\sup_{0 \leq m \leq M} E[\eta^m(t_i)] = O(n^{-\omega^2 \kappa(n)}).$$

Proof. Recall that

$$\hat{g}_{high}(t_i) = \sum_{j=0}^q \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk} I\{|\hat{d}_{jk}| > \delta_j\} \psi_{jk}(t_s).$$

The triangle inequality implies for every fixed $\omega \in (0, 1)$ (see [42])

$$\begin{aligned} I\{|d_{jk}| > (1+\omega)\delta_j\} - I\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\} &\leq I\{|\hat{d}_{jk}| > \delta_j\} \\ &\leq I\{|d_{jk}| > (1-\omega)\delta_j\} + I\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\}. \end{aligned}$$

Now, due to the definition of δ_j and [Lemma 1](#), we have for all $j \in \{0, \dots, q\}$ and n large enough,

$$I\{|d_{jk}| > (1 - \omega)\delta_j\} = I\{|d_{jk}| > (1 + \omega)\delta_j\} = I\{d_{jk} \in \bar{D}_{jk}\},$$

so that

$$A_n(t_i) - B_n(t_i) \leq \hat{g}_{high}(t_i) \leq A_n(t_i) + B_n(t_i)$$

with

$$A_n(t_i) = \sum_{j=0}^q \sum_{k=-N+1}^{N2^{j+1}-1} \hat{d}_{jk} \psi_{jk}(t_i) I\{d_{jk} \in \bar{D}_{jk}\},$$

$$B_n(t_i) = \sum_{j=0}^q \sum_{k=-N+1}^{N2^{j+1}-1} |\hat{d}_{jk} \psi_{jk}(t_i)| I\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\}$$

and

$$\hat{g}_{high}(t_i) = A_n(t_i) + \eta(t_i)$$

with

$$|\eta(t_i)| \leq B_n(t_i). \quad (25)$$

Before considering moments of $\eta(t_i)$ note that [Lemma 1](#) implies

$$\max_k E(\hat{d}_{jk}) = O(\sigma_j)$$

for $0 \leq j \leq \log_2 n - J$, and also, since $\hat{d}_{jk}$ is normal,

$$E(\hat{d}_{jk}^{2m}) = O(\sigma_j^{2m}) = O(n^{-m\alpha} 2^{-m(J+j)(1-\alpha)}).$$

Let

$$I_{jk} = I\left\{\frac{|\hat{d}_{jk} - E\hat{d}_{jk}|}{\sigma_j} > \frac{\omega\delta_j}{\sigma_j} - \frac{|E\hat{d}_{jk} - d_{jk}|}{\sigma_j}\right\}$$

and $P_{jk} = E(I_{jk})$. The triangle inequality implies $I\{|\hat{d}_{jk} - d_{jk}| > \omega\delta_j\} \leq I_{jk}$. Using this, the compact support of wavelets, [Lemmas 1\(c\)](#) and (e), the condition $\delta_j^2 = 4\kappa(n) \ln n \sigma_j^2$ ((17) and [Lemma 1\(c\)](#)) and the general inequality $P(Z > \eta) \leq e^{-\frac{1}{2}\eta^2}$ for standard normal variables Z (see e.g. [61] p. 191) leads to

$$P_{jk} \leq 2n^{-2\omega^2\kappa(n)} + o(n^{-2\omega^2\kappa(n)}).$$

Hölder's inequality together with (25) imply

$$|E[\eta(t_i)]| \leq \sum_{j=0}^q \sum_{k=-N+1}^{N2^{j+1}-1} E[|\hat{d}_{jk}| I_{jk}] |\psi_{jk}(t_i)|$$

$$\leq \sum_{j=0}^q \sum_{k=Nn^{-1}2^{j+1}-N}^{Nn^{-1}2^{j+1}-1} \sqrt{E\hat{d}_{jk}^2 \cdot P_{jk}} |\psi_{jk}(t_i)| = O(n^{-\omega^2\kappa(n)}).$$

By analogous arguments one obtains

$$|E[\eta^m(t_i)]| \leq \left(\sum_{j=0}^q \sum_{k=-N+1}^{N2^{j+1}-1} (E\hat{d}_{jk}^{2m} P_{jk})^{1/2m} |\psi_{jk}(t_i)| \right)^m = O(n^{-\omega^2\kappa(n)}). \quad \square$$

The following two lemmas provide asymptotic approximations of the first two conditional moments of $\zeta_{1,low}^*$ under the null hypothesis.

Lemma 4. Suppose that the assumptions of [Lemma 2](#) hold. Let $l = O(n^{1-\epsilon})$ for some $1 - 2(2r + \alpha)^{-1} < \epsilon < 1$. Then

$$a_l^{-1} E_n(\zeta_{1,low}^*) = o_p(1).$$

Proof.

$$a_l^{-1} E_n(\zeta_{1,low}^*) = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} X_{j,low} = S_1 + S_2$$

with

$$S_1 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} (\xi_j - Z_j),$$

$$S_2 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \{g(t_j) - E[\hat{g}_{low}(t_j)]\}$$

and $Z_j = \hat{g}_{low}(t_j) - E[\hat{g}_{low}(t_j)] = o_p(\xi_j)$ (see Lemma 2). Then $S_1 = o_p(1)$ (due to Lemma 3.1 in [53]) and Lemma 2 together with (14) imply

$$S_2 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} O(2^{-j}) = O(l a_l^{-1} 2^{-J}) = O\left(n^{-\frac{\alpha}{2r+\alpha} + \frac{\alpha(1-\epsilon)}{2}}\right)$$

which is $o(1)$ because $\epsilon > 1 - 2(2r + \alpha)^{-1}$. $\square$

Lemma 5. Suppose that the assumptions of Lemma 2 hold. Let $\hat{\sigma}_n^2 = a_l^{-2} E_n[\zeta_{1,low}^{*2}]$ and $l = O(n^{1-\epsilon})$ with $1 - 2(2r + \alpha)^{-1} < \epsilon < 1$. Then

$$\hat{\sigma}_n^2 = 2\sigma^2(1 - \alpha)^{-1}(2 - \alpha)^{-1} + o_p(1).$$

Proof.

$$\hat{\sigma}_n^2 = a_l^{-2} E_n[\zeta_{1,low}^{*2}] = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} X_{j,low} \right)^2 = S_1 + S_2 + S_3$$

with

$$S_1 = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (\xi_j - Z_j) \right)^2,$$

$$S_2 = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (g(t_j) - E[\hat{g}_{low}(t_j)]) \right)^2,$$

$$S_3 = \frac{2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (\xi_j - Z_j) \cdot \sum_{j=i}^{i+l-1} (g(t_j) - E[\hat{g}_{low}(t_j)]) \right)$$

and Z_j as before. Then

$$S_1 = \frac{2\sigma^2}{(1 - \alpha)(2 - \alpha)} + o_p(1)$$

(see Lemma 3.2 of [53]) and Lemma 2 implies

$$S_2 = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} O(2^{-j}) \right)^2 = O(l^2 a_l^{-2} 2^{-2J}) = O(n^{\alpha(1-\epsilon) - \frac{2\alpha}{2r+\alpha}})$$

which is $o(1)$ because $\epsilon > 1 - 2(2r + \alpha)^{-1}$. Finally, noting that $E(S_3) = 0$, using Hölder's inequality and writing

$$v(i, l) = \text{Var} \left(\sum_{j=i}^{i+l-1} (\xi_j - Z_j) \right), \quad w(i, l) = \sum_{j=i}^{i+l-1} \{g(t_j) - E[\hat{g}_{low}(t_j)]\}$$

we obtain

$$\text{Var}(S_3) \leq \frac{4}{a_l^4(n-l+1)^2} \max_i w^2(i, l) \sum_{i_1, i_2=1}^{n-l+1} \sqrt{v(i_1, l)v(i_2, l)} = O(n^{\alpha(1-\epsilon) - \frac{2\alpha}{2r+\alpha}}) = o(1). \quad \square$$

Now we consider $\zeta_{1,low}^{*2}$ under H_1 . Without loss of generality it is sufficient to state the results for the case with one jump.

Lemma 6. Suppose that (A1)–(A3) hold. Furthermore assume that g has a single jump of size Δ at a point $s^* \in (0, 1)$. Let $l = O(n^{1-\epsilon})$ with $1 - 2(2r + \alpha)^{-1} < \epsilon < 1$. Then

$$a_l^{-1} E_n(\zeta_{1,low}^*) = o_p(1)$$

and

$$a_l^{-2} E_n(\zeta_{1,low}^{*2}) = \frac{2\sigma^2}{(1-\alpha)(2-\alpha)} + v_n + o_p(1)$$

with

$$v_n = \Delta^2 C n^{-\frac{\alpha}{2r+\alpha} + \alpha(1-\epsilon)}$$

and $C > 0$. In particular, if $\epsilon < 1 - (2r + \alpha)^{-1}$, then $v_n \rightarrow \infty$.

Proof. Recall that

$$g(t_s) - \hat{g}_{low}(t_s) = O(2^J n^{-1}) - Z_s + \sum_{j=0}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} d_{jk} \psi_{jk}(t_s).$$

The compact support of the wavelets implies that for $k \notin [Nn^{-1}2^{J+j}s - N, Nn^{-1}2^{J+j}s]$, $\psi_{jk}(t_s)$ is equal to zero. Similarly, $\psi_{jk}(t_i) = 0$ if $i \notin [s - n2^{-(J+j)}, s + n2^{-(J+j)}]$ and $k \in [Nn^{-1}2^{J+j}s - N, Nn^{-1}2^{J+j}s]$. Let $i^* \in \mathbb{N}$ be such that $i^*/n \leq s^* \leq (i^* + 1)/n$. Then Lemma 2 implies that $g(t_s) - \hat{g}_{low}(t_s) = -Z_s + O(2^{-J})$ for $s \notin [i^* - n2^{-J}, i^* + 1 + n2^{-J}]$.

Also, for $s \in [i^* - n2^{-J}, i^* + 1 + n2^{-J}]$,

$$\sum_{j=0}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} d_{jk} \psi_{jk}(t_s) = g(t_s) - \sum_{k=-N+1}^{N2^J-1} s_{jk} \phi_{jk}(t_s) = O(1).$$

Define $f(x) = g(x) - \Delta \cdot H(x - s^*)$ where $H(x) = I\{x \geq 0\}$ is the Heaviside function, and denote by $\hat{H}_{low}(t - s^*)$ the corresponding low resolution wavelet decomposition

$$\begin{aligned} \hat{H}_{low}(t - s^*) &= \sum_{k=-N+1}^{N2^J-1} \left(\frac{1}{n} \sum_{i=1}^n H\left(\frac{i}{n} - s^*\right) \phi_{jk}\left(\frac{i}{n}\right) \right) \phi_{jk}(t) \\ &= \sum_{k=-N+1}^{N2^J-1} \frac{1}{n} \sum_{i=i^*}^n \phi_{jk}\left(\frac{i}{n}\right) \phi_{jk}(t) = \sum_{k=N2^J t - N}^{N2^J t} \int_{s^*}^1 \phi_{jk}(x) \phi_{jk}(t) dx + o(n^{-1}). \end{aligned}$$

Note that

$$\hat{f}_{low}(t) = \hat{g}_{low}(t) - \Delta \cdot \hat{H}_{low}(t - s^*)$$

and

$$\int_0^1 \hat{H}_{low}(t - s^*) dt = \int_0^1 H(t - s^*) dt + o(n^{-1}).$$

For $t > s^* + n^{-1} + 2^{-J}$ we have

$$H(t - s^*) = \hat{H}_{low}(t - s^*) + o(n^{-1})$$

and for $t < s^* - 2^{-J}$ the same equality holds exactly. Then

$$a_l^{-1} E_n(\zeta_{1,low}^*) = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} X_{j,low} = S_1 - S_2$$

with

$$S_1 = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} [\xi_j + f(t_j) - \hat{f}_{low}(t_j)],$$

$$S_2 = \frac{\Delta}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} [H(t_j - s^*) - \hat{H}_{low}(t_j - s^*)].$$

Lemma 4 implies $S_1 = o_p(1)$. Moreover,

$$\begin{aligned} S_2 &= \frac{\Delta \cdot l}{a_l(n-l+1)} \sum_{j=i^*-n2^{-J}}^{i^*+1+n2^{-J}} [H(t_j - s^*) - \hat{H}_{low}(t_j - s^*)] + o(n^{\alpha(1-\epsilon)/2-1}) \\ &= \frac{\Delta \cdot l \cdot n}{a_l(n-l+1)} \int_0^1 [H(t - s^*) - \hat{H}_{low}(t - s^*)] dt + O(la_l^{-1}n^{-1}) + o(n^{\alpha(1-\epsilon)/2-1}) \\ &= O(n^{\alpha(1-\epsilon)/2-1}). \end{aligned}$$

Taking into account Lemma 5 we have

$$\begin{aligned} a_l^{-2} E_n(\xi_{1,low}^{*2}) &= \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} X_{j,low} \right)^2 \\ &= \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (\xi_j + f(t_j) - \hat{f}_{low}(t_j)) \right)^2 + S_1 + S_2 \\ &= 2\sigma^2(1-\alpha)^{-1}(2-\alpha)^{-1} + S_1 + S_2 + o_p(1) \end{aligned} \quad (26)$$

with

$$\begin{aligned} S_1 &= -\frac{2\Delta}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left\{ \sum_{j=i}^{i+l-1} [\xi_j + f(t_j) - \hat{f}_{low}(t_j)] \sum_{j=i}^{i+l-1} [H(t_j - s^*) - \hat{H}_{low}(t_j - s^*)] \right\}, \\ S_2 &= \frac{\Delta^2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left\{ \sum_{j=i}^{i+l-1} [H(t_j - s^*) - \hat{H}_{low}(t_j - s^*)] \right\}^2. \end{aligned}$$

Now

$$\begin{aligned} \left(\frac{\Delta^2}{a_l^2(n-l+1)} \right)^{-1} S_2 &= \sum_{i=i^*-n2^{-J}-l}^{i^*+1+n2^{-J}} \left\{ \sum_{j=i}^{i+l-1} [H(t_j - s^*) - \hat{H}_{low}(t_j - s^*)] \right\}^2 + O(l^2n^{-1}) \\ &= \sum_{i=i^*-n2^{-J}-l}^{i^*+1+n2^{-J}} \left\{ n \int_{\frac{i}{n}}^{\frac{i+l}{n}} [H(x - s^*) - \hat{H}_{low}(x - s^*)] dx + o(l) \right\}^2 + O(l^2n^{-1}) \\ &= n \int_{\frac{i^*}{n}-2^{-J}-\frac{l}{n}}^{\frac{i^*}{n}+2^{-J}} \left\{ n \int_y^{y+\frac{l}{n}} [H(x - s^*) - \hat{H}_{low}(x - s^*)] dx \right\}^2 dy + o(l^2n2^{-J}). \end{aligned}$$

Using the representation of $\hat{H}_{low}$ and $\int_y^{y+\frac{l}{n}} \phi_{jk}(x) dx = \frac{l}{n} \phi_{jk}(y) + o(\ln^{-1} 2^{J/2})$, the integral can be written as

$$\begin{aligned} n \int_{s^*-2^{-J}-\frac{l}{n}}^{s^*+2^{-J}} \left\{ n \int_y^{y+\frac{l}{n}} \left[H(x - s^*) - \sum_{k=-N+1}^{N2^J-1} \int_{s^*}^1 \phi_{jk}(t) \phi_{jk}(x) dt \right] dx \right\}^2 dy \\ = nl^2 \int_{s^*-2^{-J}-\frac{l}{n}}^{s^*+2^{-J}} \left(\tilde{H}(y) - \sum_{k=-N+1}^{N2^J-1} \int_{s^*}^1 \phi_{jk}(t) \phi_{jk}(y) dt \right)^2 dy + o(l^2n2^{-J}) \end{aligned}$$

with

$$\tilde{H}(t) = \begin{cases} 0, & y < s^* - \ln^{-1} \\ l^{-1}ny + 1 - l^{-1}s^*, & s^* - \ln^{-1} \leq y < s^* \\ 1, & y \geq s^*. \end{cases}$$

Noting that $\tilde{H}(y) - \sum_{k=-N+1}^{N2^J-1} \int_{s^*}^1 \phi_{jk}(t) \phi_{jk}(y) dt = 0$ for $y \notin [s^* - 2^{-J} - \frac{1}{n}, s^* + 2^{-J}]$ and using the wavelet decomposition of $\tilde{H}$ and orthonormality of ϕ_{jk} and ψ_{jk} we obtain

$$\begin{aligned} & n l^2 \int_0^1 \left(\tilde{H}(y) - H(y - s^*) + \sum_{j=0}^q \sum_{k=-N+1}^{N2^{J+j}-1} \int_0^1 H(t - s^*) \psi_{jk}(t) \psi_{jk}(y) dt \right)^2 dy + o(l^2 n 2^{-J}) \\ &= I_1 + I_2 + I_3 + o(l^2 n 2^{-J}) \end{aligned}$$

with

$$\begin{aligned} I_1 &= n l^2 \int_0^1 [\tilde{H}(y) - H(y - s^*)]^2 dy, \\ I_2 &= n l^2 \sum_{j=0}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} \left[\int_0^1 H(t - s^*) \psi_{jk}(t) dt \right]^2, \\ I_3 &= n l^2 \int_0^1 [\tilde{H}(y) - H(y - s^*)] \sum_{j=0}^q \sum_{k=-N+1}^{N2^{J+j}-1} \int_0^1 H(t - s^*) \psi_{jk}(t) \psi_{jk}(y) dt dy \end{aligned}$$

where

$$\begin{aligned} I_1 &= \frac{1}{3} l^3, \\ I_3 &= n l^2 \int_0^1 [\tilde{H}(y) - H(y - s^*)][H(y - s^*) - \hat{H}_{low}(y - s^*)] dy = O(l^3). \end{aligned}$$

Since $\int_0^1 H(t - s^*) \psi_{jk}(t) dt = 0$ for $\forall k \notin U_j = [\lfloor N2^{J+j} s^* \rfloor - N, \lfloor N2^{J+j} s^* \rfloor + 1]$ and $N2^{J+j} - \lfloor N2^{J+j} s^* \rfloor - k \rightarrow \infty$, we obtain

$$\begin{aligned} I_2 &= n l^2 \sum_{j=0}^{\infty} \sum_{k \in U_j} \left(\int_{s^*}^1 \psi_{jk}(t) dt \right)^2 \\ &= N^{-1} n l^2 \sum_{j=0}^{\infty} 2^{-(J+j)} \sum_{k=-N}^1 \left(\int_{N2^{J+j} s^* - \lfloor N2^{J+j} s^* \rfloor - k}^N \psi(t) dt \right)^2 \\ &\geq 2N^{-1} n l^2 2^{-J} \min_{z \in (0, 1)} \left[\sum_{k=-N}^1 \left(\int_{z-k}^N \psi(t) dt \right)^2 \right] \end{aligned}$$

and the bounds $I_{lower} \leq I_2 \leq I_{upper}$, where

$$\begin{aligned} I_{lower} &= 2N^{-1} n l^2 2^{-J} \min_{z \in (0, 1)} \left[\sum_{k=-N}^1 \left(\int_{z-k}^N \psi(t) dt \right)^2 \right] + o(l^2 n 2^{-J}), \\ I_{upper} &= N^{-1} n l^2 \sum_{j=0}^{\infty} 2^{-(J+j)} \sum_{k=-N}^1 \int_0^N \psi^2(t) dt = 2N^{-1} n l^2 2^{-J} (N + 2). \end{aligned}$$

Thus, using the definitions of l and J , there exists a constant C such that

$$S_2 = \Delta^2 C n^{-\frac{\alpha}{2r+\alpha} + \alpha(1-\epsilon)} + o(n^{-\frac{\alpha}{2r+\alpha} + \alpha(1-\epsilon)}).$$

Finally, by similar detailed calculations of $E(S_1)$ and $\text{Var}(S_1)$ it can be shown that S_1 is asymptotically negligible. $\square$

The following two lemmas compare the conditional expected values and conditional variances of ζ_1^* and $\zeta_{1,low}^*$ under H_0 .

Lemma 7. Suppose that the assumptions of Lemma 2 hold. Let $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$. Then

$$a_l^{-1} E_n(\zeta_1^*) = a_l^{-1} E_n(\zeta_{1,low}^*) + O_p(n^{\frac{\alpha}{2}(1-\epsilon) - \frac{1}{2}\omega^2 \kappa(n)}).$$

Proof.

$$a_l^{-1} E_n(\zeta_1^*) = a_l^{-1} E_n(\zeta_{1,low}^*) - S$$

with

$$S = \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \hat{g}_{high}(t_j).$$

From Lemma 3 we have

$$\begin{aligned} E(S) &= \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{s=i}^{i+l-1} \left(\sum_{j=0}^q \sum_{k=-N+1}^{N2^{l+j}-1} E[\hat{d}_{jk}] I(d_{jk} \in \bar{D}_j) \psi_{jk}(t_s) + E[\eta(t_s)] \right) \\ &= \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{s=i}^{i+l-1} E[\eta(t_s)] = O(n^{\frac{\alpha}{2}(1-\epsilon) - \omega^2 \kappa(n)}) \end{aligned}$$

and

$$\begin{aligned} E(S^2) &= \frac{1}{a_l^2(n-l+1)^2} E \left[\left(\sum_{i=1}^{n-l+1} \sum_{s=i}^{i+l-1} \eta(t_s) \right)^2 \right] \\ &\leq \frac{1}{a_l^2(n-l+1)^2} \sum_{i_1=1}^{n-l+1} \sum_{s_1=i_1}^{i_1+l-1} \sum_{i_2=1}^{n-l+1} \sum_{s_2=i_2}^{i_2+l-1} \sqrt{E[\eta^2(t_{s_1})] E[\eta^2(t_{s_2})]} = O(n^{\alpha(1-\epsilon) - \omega^2 \kappa(n)}). \quad \square \end{aligned}$$

Lemma 8. Suppose that the assumptions of Lemma 2 hold. Let $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$. Then

$$a_l^{-2} E_n(\zeta_1^{*2}) = a_l^{-2} E_n(\zeta_{1,low}^{*2}) + O_p(a_l^{-2} l \cdot n^{-1/2} n^{-\frac{1}{4} \omega^2 \kappa(n)}).$$

Proof.

$$\begin{aligned} \hat{\sigma}_n^2 &= a_l^{-2} E_n(\zeta_1^{*2}) = \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} X_j \right)^2 \\ &= a_l^{-2} E_n(\zeta_{1,low}^{*2}) - S_1 + S_2 \end{aligned}$$

with

$$\begin{aligned} S_1 &= \frac{2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} (\xi_j + g(t_j) - \hat{g}_{low}(t_j)) \sum_{j=i}^{i+l-1} \hat{g}_{high}(t_j) \right), \\ S_2 &= \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \left(\sum_{j=i}^{i+l-1} \hat{g}_{high}(t_j) \right)^2. \end{aligned}$$

Now

$$E(S_2) \leq \frac{1}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} E[\hat{g}_{high}^2(t_j)] = O(l a_l^{-2} n^{-\omega^2 \kappa(n)})$$

and

$$E(S_2^2) \leq \frac{1}{a_l^4(n-l+1)^2} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} E[\hat{g}_{high}^4(t_j)] = O(l a_l^{-4} n^{-1} n^{-\omega^2 \kappa(n)}).$$

Furthermore,

$$\begin{aligned} |E(S_1)| &\leq \frac{2}{a_l^2(n-l+1)} \sum_{i=1}^{n-l+1} \sqrt{E \left[\left(\sum_{j=i}^{i+l-1} [\xi_j + g(t_j) - \hat{g}_{low}(t_j)] \right)^2 \right]} \sum_{j=i}^{i+l-1} \sqrt{E[\hat{g}_{high}^2(t_j)]} \\ &= O(l^{2-\alpha/2} a_l^{-2} n^{-\frac{1}{2} \omega^2 \kappa(n)}) \end{aligned}$$

and

$$\begin{aligned} E(S_1^2) &\leq \frac{4}{a_l^4(n-l+1)^2} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} \sqrt{E[(\xi_j + g(t_j) - \hat{g}_{low}(t_j))^4]} \sum_{j'=i}^{i+l-1} \sqrt{E[\hat{g}_{high}^4(t_{j'})]} \\ &= O(a_l^{-4} l^2 n^{-1} n^{-\frac{1}{2}\omega^2 \kappa(n)}). \quad \square \end{aligned}$$

Finally, we derive asymptotic expressions for the first two conditional moments of ζ_1^* under H_1 .

Lemma 9. Suppose that the assumptions of Lemma 6 hold. Let $l = O(n^{1-\epsilon})$ with $1 - 2(2r + \alpha)^{-1} < \epsilon < 1$. Then

$$a_l^{-1} E_n(\zeta_1^*) = o_p(1)$$

and

$$a_l^{-2} E_n(\zeta_1^{*2}) = \frac{2\sigma^2}{(1-\alpha)(2-\alpha)} + o_p(1).$$

Proof. We use the same notation as in the proof of Lemma 6. Recall that

$$\hat{g}_{high}(t_s) = \sum_{j=0}^q \sum_{k=-N+1}^{N2^{l+j}-1} \hat{d}_{jk} I(d_{jk} \in \bar{D}_j) \psi_{jk}(t_s) + \eta(t_s),$$

for $k \notin [N2^{l+j}n^{-1}i^* - N, N2^{l+j}n^{-1}i^*]$ we have $\psi_{jk}(s^*) = 0$ and $d_{jk} \in D_j$ and also for $k \in [N2^{l+j}s^* - N, N2^{l+j}s^*]$ and t outside the interval $[s^* - 2^{-(l+j)}, s^* + 2^{-(l+j)}]$, one has $\psi_{jk}(t) = 0$. Defining $U = \{i^* - n2^{-l} - l, \dots, i^* + n2^{-l}\}$ it then follows that

$$\begin{aligned} a_l^{-1} E_n(\zeta_1^*) &= \frac{1}{a_l(n-l+1)} \sum_{i=1}^{n-l+1} \sum_{j=i}^{i+l-1} (\xi_j + g(t_j) - \hat{g}_{low}(t_j) - \hat{g}_{high}(t_j)) \\ &= \frac{1}{a_l(n-l+1)} \sum_{i \notin U} \sum_{j=i}^{i+l-1} + \frac{1}{a_l(n-l+1)} \sum_{i \in U} \sum_{j=i}^{i+l-1} = S_1 + S_2. \end{aligned}$$

For S_1 the result is as before (since $\bar{U}$ does not include jumps). For S_2 consider first

$$g(t_s) - \hat{g}_{low}(t_s) - \hat{g}_{high}(t_s) = A_{1,s} + A_{2,s} + A_{3,s} + A_{4,s}$$

with

$$\begin{aligned} A_{1,s} &= \sum_{k=-N+1}^{N2^l-1} (s_{jk} - \hat{s}_{jk}) \phi_{jk}(t_s), & A_{2,s} &= \sum_{j=0}^q \sum_{k=-N+1}^{N2^{l+j}-1} d_{jk} I\{d_{jk} \in D_j\} \psi_{jk}(t_s), \\ A_{3,s} &= \sum_{j=0}^q \sum_{k=-N+1}^{N2^{l+j}-1} (d_{jk} - \hat{d}_{jk}) I\{d_{jk} \in \bar{D}_j\} \psi_{jk}(t_s), & A_{4,s} &= \sum_{j=q+1}^{\infty} \sum_{k=-N+1}^{N2^{l+j}-1} d_{jk} \psi_{jk}(t_s). \end{aligned}$$

If $s \in [i^* - 1, i^* + 1]$, then

$$A_{4,s} = g(t_s) - \sum_{k=-N+1}^{N2^{l+q}-1} s_{(j+q)k} \phi_{(j+q)k}(t_s) = O(1).$$

On the other hand, if $s \notin [i^* - 1, i^* + 1]$, then $A_{4,s} = O(n^{-1})$. Furthermore $A_{2,s} = O(2^{-j})$. Finally, for $a \in \{0, 1, \dots, \log_2 n - j\}$, and s between $i^* - n2^{-(l+a)}$ and $i^* + n2^{-(l+a)}$ but not in the interval $[i^* - n2^{-(l+a+1)}, i^* + n2^{-(l+a+1)}]$, we have

$$\begin{aligned} E(A_{3,s}) &= \sum_{j=0}^q \sum_{k=-N+1}^{N2^{l+j}-1} [d_{jk} - E(\hat{d}_{jk})] I\{d_{jk} \in \bar{D}_j\} \psi_{jk}(t_s) \\ &= \sum_{j=0}^a + \sum_{j=a+1}^q = O(n^{-1} 2^{l+a}) \end{aligned}$$

and

$$\begin{aligned} \text{Var}(A_{3,s}) &= E \left\{ \left(\sum_{j=0}^a \sum_{k=-N+1}^{N2^{j+1}-1} [E(\hat{d}_{jk}) - \hat{d}_{jk}] I\{d_{jk} \in \bar{D}_j\} \psi_{jk}(t_s) \right)^2 \right\} + O(n^{-2} 2^{2(j+a)}) \\ &\leq \left(\sum_{j=0}^a \sum_{k=-N+1}^{N2^{j+1}-1} \sigma_j |\psi_{jk}(t_s)| I(d_{jk} \in \bar{D}_j) \right)^2 + O(n^{-2} 2^{2(j+a)}) = O(n^{-\alpha} 2^{\alpha(j+a)}). \end{aligned}$$

Note that $A_{3,s}$ is normally distributed so that also $E(A_{3,s}^4) = O(n^{-2\alpha} 2^{2\alpha(j+a)})$. For S_2 we then obtain

$$\begin{aligned} |E(S_2)| &= \left| E \left[\frac{1}{a_l(n-l+1)} \sum_{i \in U} \sum_{j=i}^{i+l-1} (\xi_j + Z_j + A_{3,j} + O(2^{-j})) \right] \right| + O(la_l^{-1}n^{-1}) \\ &\leq \frac{l}{a_l(n-l+1)} \sum_{j=s^*-n2^{-j}}^{s^*+n2^{-j}} |E(A_{3,j})| + o(1) \\ &\leq \frac{l}{a_l(n-l+1)} \sum_{a=0}^q (n2^{-(j+a)} - n2^{-(j+a+1)}) O(n^{-1} 2^{j+a}) = O(qla_l^{-1}n^{-1}) = o(1). \end{aligned}$$

Similarly one obtains

$$E(S_2^2) = S_{21} + S_{22} + o(1)$$

with

$$\begin{aligned} S_{21} &= \frac{2}{a_l^2(n-l+1)^2} E \left[\sum_{i \in U} \sum_{j=i}^{i+l-1} (\xi_j + Z_j) \sum_{i \in U} \sum_{j=i}^{i+l-1} A_{3,j} \right] \\ &= O(l^2 a_l^{-2} n^{-\alpha} 2^{-(2-\alpha)l}) = o(1), \\ S_{22} &= \frac{1}{a_l^2(n-l+1)^2} E \left[\left(\sum_{i \in U} \sum_{j=i}^{i+l-1} A_{3,j} \right)^2 \right] \\ &= o(la_l^{-1} n^{-\alpha/2} 2^{-(1-\alpha/2)l}) = o(1). \end{aligned}$$

Thus, the result for $a_l^{-1} E_n(\zeta_1^*)$ follows.

For the second moment we have

$$\sigma_n^2 = a_l^{-2} E_n(\zeta_1^{*2}) = \frac{2\sigma^2}{(1-\alpha)(2-\alpha)} + S_3 + o_p(1)$$

with

$$\begin{aligned} S_3 &= \frac{1}{a_l^2(n-l+1)} \sum_{i \in U} \left\{ \sum_{j=i}^{i+l-1} [\xi_j + g(t_j) - \hat{g}_{low}(t_j) - \hat{g}_{high}(t_j)] \right\}^2 \\ &= \frac{1}{a_l^2(n-l+1)} (S_{31} + S_{32} + S_{33} + S_{34} + S_{35} + S_{36}) + o(1) \end{aligned}$$

and

$$\begin{aligned} S_{31} &= \sum_{i \in U} \left(\sum_{j=i}^{i+l-1} (\xi_j + Z_j) \right)^2, \quad S_{32} = \sum_{i \in U} \left(\sum_{j=i}^{i+l-1} A_{3,j} \right)^2, \quad S_{33} = \sum_{i \in U} \left(\sum_{j=i}^{i+l-1} O(2^{-j}) \right)^2, \\ S_{34} &= 2 \sum_{i \in U} \left[\left(\sum_{j=i}^{i+l-1} (\xi_j + Z_j) \right) \left(\sum_{j=i}^{i+l-1} A_{3,j} \right) \right], \\ S_{35} &= 2 \sum_{i \in U} \left[\left(\sum_{j=i}^{i+l-1} (\xi_j + Z_j) \right) \left(\sum_{j=i}^{i+l-1} O(2^{-j}) \right) \right], \end{aligned}$$

$$S_{36} = \sum_{i \in U} \left[\left(\sum_{j=i}^{i+l-1} \gamma_j \right) \left(\sum_{j=i}^{i+l-1} o(2^{-j}) \right) \right].$$

A detailed analysis of the six terms shows that they are asymptotically negligible so that $S_3 = o_p(1)$. $\square$

The results of [Lemmas 6](#) and [9](#) remain the same in the presence of any finite number of isolated jumps.

Proof of Theorem 1. The results follow directly from [Lemmas 4–6](#) and definition [\(19\)](#). $\square$

Proof of Theorem 2. The results follow from [Theorem 1](#) in an analogous manner as [Theorem 2.2](#) in [\[53\]](#). $\square$

Proof of Theorem 3. The theorem follows directly from [Lemmas 7–9](#) with $\omega = 2^{-1/2}$ and the definitions [\(18\)](#), [\(19\)](#). $\square$

Proof of Theorem 4. This follows from [Theorem 3](#) in a similar way as [Theorem 2](#). $\square$

References

- [1] F. Abramovich, T. Sapatinas, B.W. Silverman, Wavelet thresholding via a Bayesian approach, *Journal of Royal Statistical Society* 60 (4) (1998) 725–749.
- [2] Y. Ait-Sahalia, J. Jacod, Estimating the degree of activity of jumps in high frequency data, *The Annals of Statistics* 37 (5A) (2009) 2202–2244.
- [3] D.W.K. Andrews, O. Lieberman, V. Marmer, Higher-order improvements of the parametric bootstrap for long-memory Gaussian processes, *Journal of Econometrics* 133 (2) (2006) 673–702.
- [4] J. Arteche, J. Orbe, Bootstrapping the log-periodogram regression, *Economics Letters* 86 (1) (2005) 79–85.
- [5] A. Banerjee, G. Urga, Modelling structural breaks, long memory and stock market volatility: an overview, *Journal of Econometrics* 129 (1–2) (2005) 1–34.
- [6] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, London, 1994.
- [7] J. Beran, Fitting long memory models by generalized regression, *Biometrika* 80 (4) (1993) 817–822.
- [8] J. Beran, Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models, *JRSS, Series B* 57 (4) (1995) 659–672.
- [9] J. Beran, Y. Feng, SEMIFAR models – a semiparametric framework for modelling trends, long-range dependence and nonstationarity, *Computational Statistics & Data Analysis* 40 (2) (2002) 393–419.
- [10] J. Beran, Y. Feng, Data driven bandwidth choice for SEMIFAR models, *Journal of Computational and Graphical Statistics* 11 (2) (2002) 690–713.
- [11] J. Beran, Y. Feng, Local polynomial fitting with long memory, short memory and antipersistent errors, *Annals of the Institute of Statistical Mathematics* 54 (2) (2002) 291–311.
- [12] J. Beran, Y. Feng, S. Ghosh, P. Sibbertsen, On robust local polynomial estimation with long-memory errors, *International Journal of Forecasting* 18 (2) (2002) 227–241.
- [13] J. Beran, Y. Shumeyko, On asymptotically optimal wavelet estimation of trend functions under long-range dependence, *Bernoulli*, (2011) (in press).
- [14] I. Berkes, L. Horváth, P. Kokoszka, Q. Shao, On discriminating between long-range dependence and changes in mean, *The Annals of Statistics* 34 (3) (2006) 1140–1165.
- [15] D. Brillinger, Uses of cumulants in wavelet analysis, *Proceedings of the SPIE, Advanced Signal Processing* 2296 (1994) 2–18.
- [16] D. Brillinger, Some uses of cumulants in wavelet analysis, *Nonparametric Statistics* 6 (1996) 93–114.
- [17] K. Choi, E. Zivot, Long memory and structural changes in the forward discount: an empirical investigation, *Journal of International Money and Finance* 26 (3) (2007) 342–363.
- [18] A. Cohen, I. Daubechies, P. Vial, Wavelets on the interval and fast wavelet transforms, *Applied and Computational Harmonic Analysis* 1 (1) (1993) 54–81.
- [19] M. Csörgő, L. Horváth, *Limit Theorems in Change-Point Analysis*, Wiley, New York, 1998.
- [20] S. Csörgő, J. Miśkiewicz, Nonparametric regression under long-range dependent normal errors, *The Annals of Statistics* 23 (3) (1995) 1000–1014.
- [21] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- [22] J. Davidson, A model of fractional cointegration, and tests for cointegration using the bootstrap, *Journal of Econometrics* 110 (2) (2002) 187–212.
- [23] M.A. Delgado, J. Hidalgo, Nonparametric inference on structural breaks, *Journal of Econometrics* 96 (1) (2000) 113–144.
- [24] F.X. Diebold, A. Inoue, Long memory and regime switching, *Journal of Econometrics* 105 (1) (2001) 131–159.
- [25] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika* 81 (3) (1994) 425–455.
- [26] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association* 90 (432) (1995) 1200–1224.
- [27] D.L. Donoho, I.M. Johnstone, Minimax estimation via wavelet shrinkage, *The Annals of Statistics* 26 (3) (1998) 879–921.
- [28] R. Fox, M.S. Taqqu, Large sample properties of parameter estimates for strongly dependent stationary time series, *The Annals of Statistics* 14 (2) (1986) 517–532.
- [29] G.C. Franco, V.A. Reisen, Bootstrap techniques in semiparametric estimation methods for ARFIMA models: a comparison study, *Computational Statistics* 19 (2) (2004) 243–259.
- [30] T. Gasser, H.G. Müller, Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics* 11 (3) (1984) 171–185.
- [31] J. Geweke, S. Porter-Hudak, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4 (4) (1983) 221–238.
- [32] S. Ghosh, J. Beran, J. Innes, Nonparametric conditional quantile estimation in the presence of long memory, *Student* 2 (1997) 109–117.
- [33] S. Ghosh, D. Draghicescu, Predicting the distribution function for long-memory processes, *International Journal of Forecasting* 18 (2) (2002) 283–290.
- [34] S. Ghosh, D. Draghicescu, An algorithm for optimal bandwidth selection for smooth nonparametric quantiles and distribution functions, in: *Statistics in Industry and Technology: Statistical Data Analysis Based on the L1-Norm and Related Methods*, Birkhäuser Verlag, Basel, Switzerland, 2002, pp. 161–168.
- [35] L.A. Gil-Alana, Fractional integration and structural breaks at unknown periods of time, *Journal of Time Series Analysis* 29 (1) (2008) 163–185.
- [36] L. Giraitis, D. Surgailis, A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle’s estimate, *Probability Theory Related Fields* 86 (1) (1990) 87–104.
- [37] C.W.J. Granger, N. Hyung, Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance* 11 (3) (2004) 399–421.
- [38] P. Hackl, A. Westlund, *Economic Structural Change, Analysis and Forecasting*, Springer, Berlin, 1991.
- [39] P. Hall, W. Härdle, T. Kleinow, P. Schmidt, Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the Hurst coefficient, *Statistical Inference for Stochastic Processes* 3 (3) (2000) 263–276.
- [40] P. Hall, J.D. Hart, Nonparametric regression with long-range dependence, *Stochastic Processes and Their Applications* 36 (2) (1990) 339–351.
- [41] P. Hall, B.Y. Jing, S.N. Lahiri, On the sampling window method for long-range dependent data, *Statistica Sinica* 8 (4) (1998) 1189–1204.
- [42] P. Hall, P. Patil, Formulae for mean integrated squared error of non-linear wavelet-based density estimators, *The Annals of Statistics* 23 (3) (1995) 905–928.

- [43] P. Hall, P. Patil, On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods, *Journal of the Royal Statistical Society, Series B* 58 (2) (1996) 361–377.
- [44] P. Hall, P. Patil, Effect of threshold rules on performance of wavelet-based curve estimators, *Statistica Sinica* 6 (2) (1996) 331–345.
- [45] J. Hidalgo, P.M. Robinson, Testing for structural change in a long memory environment, *Journal of Econometrics* 70 (1) (1996) 159–174.
- [46] L. Horváth, Change-point detection in long-memory processes, *Journal of Multivariate Analysis* 78 (2) (2001) 218–234.
- [47] L. Horváth, P. Kokoszka, The effect of long-range dependence on change-point estimators, *Journal of Statistical Planning and Inference* 64 (1) (1997) 57–81.
- [48] C. Hurvich, R. Deo, J. Brodsky, The mean squared error of Geweke and Porter-Hudak's estimator of a long-memory time-series, *Journal of Time Series Analysis* 19 (1) (1998) 19–46.
- [49] I.M. Johnstone, Wavelet threshold estimators for correlated data and inverse problems: adaptivity results, *Statistica Sinica* 9 (1) (1999) 51–83.
- [50] I.M. Johnstone, B.W. Silverman, Wavelet threshold estimators for data with correlated noise, *Journal of the Royal Statistical Society, Series B* 59 (2) (1997) 319–351.
- [51] W. Krämer, P. Sibbertsen, Testing for structural changes in the presence of long memory, *International Journal of Business and Economics* 1 (3) (2002) 235–242.
- [52] R. Kulik, M. Raimondo, L^p -wavelet regression with correlated errors and inverse problems, *Statistica Sinica* 19 (4) (2009) 1479–1489.
- [53] S.N. Lahiri, On the moving block bootstrap under long range dependence, *Statistics & Probability Letters* 18 (5) (1993) 405–413.
- [54] S. Lazarová, Testing for structural change in regression with long memory processes, *Journal of Econometrics* 129 (1–2) (2005) 329–372.
- [55] L. Li, Y. Xiao, Mean integrated squared error of nonlinear wavelet-based estimators with long memory data, *Annals of the Institute of Statistical Mathematics* 59 (2) (2007) 299–324.
- [56] T. Mikosch, C. Stărică, Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects, *The Review of Economics and Statistics* 86 (1) (2004) 378–390.
- [57] E. Moulines, P. Soulier, Broadband log-periodogram regression of time series with long-range dependence, *The Annals Statistics* 27 (4) (1999) 1415–1439.
- [58] G.P. Nason, Wavelet shrinkage using cross-validation, *Journal of the Royal Statistical Society, Series B* 58 (2) (1996) 463–479.
- [59] D.B. Percival, S. Sardy, A.C. Davison, Wavestrapping time series: adaptive wavelet-based bootstrapping, in: W.J. Fitzgerald, R.L. Smith, A.T. Walden (Eds.), *Nonlinear and Nonstationary Signal Processing*, Cambridge University Press, Cambridge, 2000, pp. 442–471.
- [60] D.P. Percival, A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, 2000.
- [61] D. Pollard, *Convergence of Stochastic Processes*, Springer-Verlag, New York Inc., 1984.
- [62] D.S. Poskitt, Properties of the sieve bootstrap for fractionally integrated and non-invertible processes, *Journal of Time Series Analysis* 29 (2) (2008) 224–250.
- [63] B.K. Ray, R.S. Tsay, Bayesian methods for change-point detection in long-range dependent processes, *Journal of Time Series Analysis* 23 (6) (2002) 687–705.
- [64] B.K. Ray, R.S. Tsay, Bandwidth selection for kernel regression with long-range dependent errors, *Biometrika* 84 (4) (1997) 791–802.
- [65] P.M. Robinson, Large sample inference for nonparametric regression with dependent errors, *The Annals of Statistics* 25 (5) (1997) 2054–2083.
- [66] P.M. Robinson, Log-periodogram regression of time series with long range dependence, *The Annals Statistics* 23 (3) (1995) 1048–1072.
- [67] P.M. Robinson, Gaussian semiparametric estimation of long range dependence, *The Annals Statistics* 23 (5) (1995) 1630–1661.
- [68] G. Samorodnitsky, *Long-Range Dependence*, Now Publishers Inc., Hanover, USA, 2007.
- [69] P. Sibbertsen, Long memory versus structural breaks: an overview, *Statistical Papers* 45 (4) (2004) 465–515.
- [70] E.M. Silva, G.C. Franco, V.A. Reisen, F.R.B. Cruz, Local bootstrap approaches for fractional differential parameter estimation in ARFIMA models, *Computational Statistics and Data Analysis* 51 (2) (2006) 1002–1011.
- [71] Y.K. Truong, P.N. Patil, Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series, *The Annals of Institute of Statistical Mathematics* 53 (1) (2001) 159–178.
- [72] B. Vidakovic, *Statistical Modeling by Wavelets*, Wiley, New York, 1999.
- [73] R. von Sachs, B. Macgibbon, Non-parametric curve estimation by wavelet thresholding with locally stationary errors, *Scandinavian Journal of Statistics* 27 (3) (2000) 475–499.
- [74] Y. Wang, Function estimation via wavelet shrinkage for long-memory data, *The Annals of Statistics* 24 (2) (1996) 466–484.
- [75] J.H. Wright, Testing for a structural break at unknown date with long-memory disturbances, *Journal of Time Series Analysis* 19 (3) (1998) 369–376.
- [76] Y. Yang, Nonparametric regression with dependent errors, *Bernoulli* 7 (4) (2001) 633–655.
- [77] V.A. Zorich, *Mathematical Analysis*, vol. 1, Springer-Verlag, Berlin, 2004.
- [78] A. Zygmund, *Trigonometric Series*, Cambridge University Press, London, 1953.