Estimation of the long-run variance of nonlinear time series with an application to change point analysis

Vaidotas Characiejus[a]

Joint work with Piotr Kokoszka[b] and Xiangdong Meng[b]

Two-Day Meeting of the Danish Statistical Society / November 4, 2025

[a] Department of Mathematics and Computer Science, University of Southern Denmark, Denmark

[b] Department of Statistics, Colorado State University, USA

Outline

Motivation

Smoothed periodogram estimator of the long-run variance

Weakly dependent nonlinear processes

Limiting distribution of the estimator

Application to change point analysis

Summary

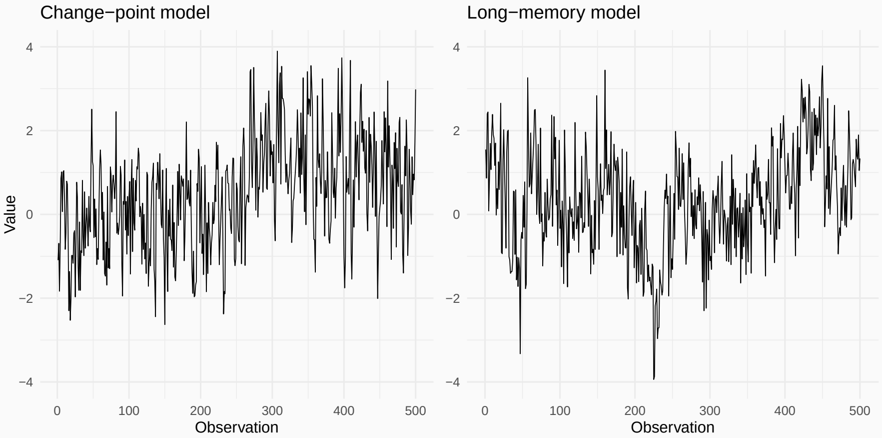
Motivation

Overview

- We focus on the smoothed periodogram estimator for the long-run variance (LRV).
- Our interest in this specific LRV estimator stems from the challenges of distinguishing between change-point and long-memory models.
- Although change-point and long-memory models are conceptually distinct, they often produce similar patterns in:
    - time plots,
    - correlograms,
    - periodograms.

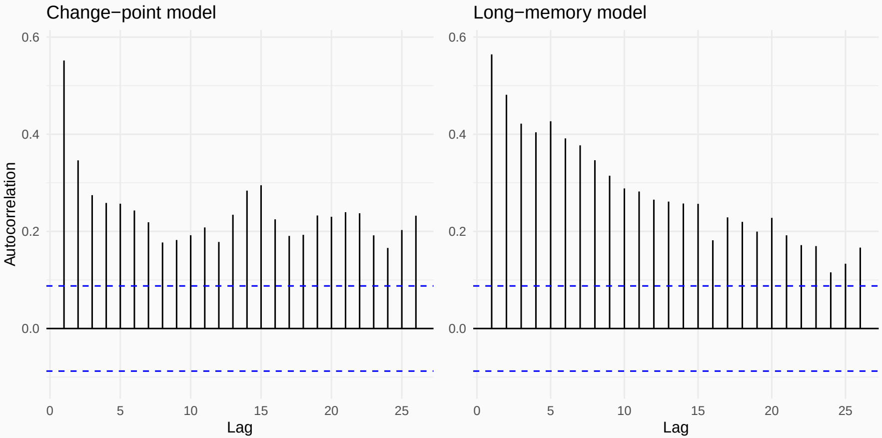
  As a result, differentiating between them in practice can be quite challenging.

Time plots

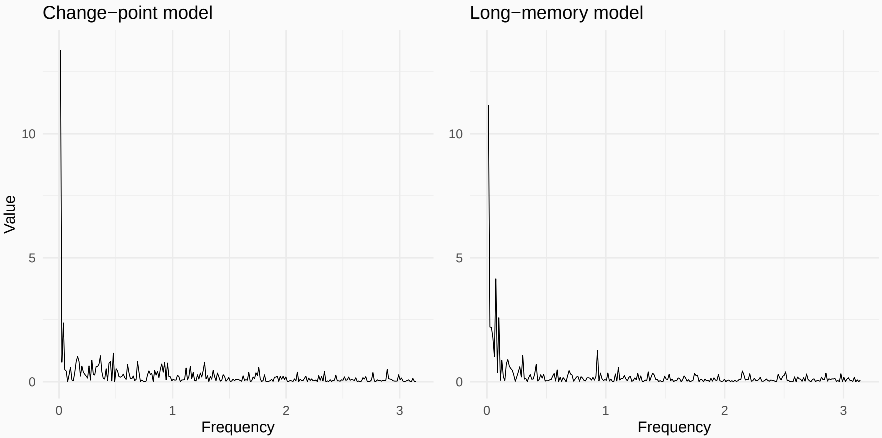


Change−point model
Long−memory model
Value
4
2
0
−2
−4
0
100
200
300
400
500
Observation

Correlograms



Periodograms



Smoothed periodogram estimator of the long-run variance

Stationary sequences and their autocovariances

Definition

$\{X_t\}_{t\in\mathbb{Z}}$ is a stationary sequence of random variables if

(a) $\mathsf{E}\,|X_t|^2 < \infty$ for all $t \in \mathbb{Z}$;

(b) $\mathsf{E}\,X_t = \mu$ for all $t \in \mathbb{Z}$;

(c) $\mathsf{Cov}(X_r, X_s) = \mathsf{Cov}(X_{r+t}, X_{s+t})$ for all $r, s, t \in \mathbb{Z}$.

Definition

For a stationary sequence $\{X_t\}_{t\in\mathbb{Z}}$,

$$\gamma(h) := \mathsf{Cov}(X_h, X_0) \quad \text{with} \quad h \in \mathbb{Z}$$

is the sequence of autocovariances.

Long-run variance

Definition

For a stationary sequence $\{X_t\}_{t\in\mathbb{Z}}$ with $\sum_{h=1}^{\infty}|\gamma(h)| < \infty$,

$$\sigma_\infty^2 := \sum_{h\in\mathbb{Z}} \gamma(h) = \gamma(0) + 2\sum_{h=1}^{\infty}\gamma(h)$$

is the long-run variance (LRV) of $\{X_t\}_{t\in\mathbb{Z}}$.

- For an uncorrelated sequence $\{X_t\}_{t\in\mathbb{Z}}$,

$$\operatorname{Var}\bar{X}_n = \frac{\gamma(0)}{n} \quad \text{for} \quad n \geq 1.$$

- For a stationary sequence $\{X_t\}_{t\in\mathbb{Z}}$ with $\sum_{h=1}^{\infty}|\gamma(h)| < \infty$,

$$\operatorname{Var}\bar{X}_n \sim \frac{\sigma_\infty^2}{n} \quad \text{as} \quad n \to \infty.$$

Spectral density

The LRV is typically estimated nonparametrically in the frequency domain.

Definition

For a stationary sequence $\{X_t\}_{t\in\mathbb{Z}}$ with $\sum_{h=1}^{\infty}|\gamma(h)| < \infty$,

$$f(\lambda) := \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma(h)e^{-ih\lambda} \quad \text{for} \quad \lambda \in [-\pi,\pi]$$

is the spectral density of $\{X_t\}_{t\in\mathbb{Z}}$, where $i = \sqrt{-1}$.

Observe that

$$\sigma_\infty^2 = 2\pi f(0)$$

and hence we can estimate the LRV by estimating $f(0)$.

Discrete Fourier transform and the periodogram

Definition

The discrete Fourier transform (DFT) and the periodogram of $X_1, \ldots, X_n$ with $n \geq 1$ are defined as

$$\mathcal{X}_n(\omega_j) := \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{-it\omega_j} \quad \text{and} \quad I_n(\omega_j) := |\mathcal{X}_n(\omega_j)|^2,$$

respectively, where $\omega_j = 2\pi j/n$ with $j \in \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}$ are the Fourier frequencies and $i = \sqrt{-1}$.

Periodogram as the sample spectral density

The periodogram can be viewed as the sample spectral density since

$$I_n(\omega_j) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-ih\omega_j},$$

where $\hat{\gamma}(h)$ with $|h| < n$ are the sample autocovariances given by

$$\hat{\gamma}(h) := n^{-1} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n)$$

for $h \geq 0$ and $\hat{\gamma}(h) = \hat{\gamma}(-h)$ for $h < 0$ with $\bar{X}_n = n^{-1} \sum_{t=1}^{n} X_t$.

Smoothed periodogram estimator

- If $\mu = 0$, the periodogram is an asymptotically unbiased estimator of $f(\lambda)$ for $\lambda \in [-\pi, \pi]$.
- The periodogram is not a consistent estimator of the spectral density.
- A simple consistent estimator of the spectral density function is constructed by smoothing the periodogram in the following way

$$\hat{f}(\lambda) = \frac{1}{2m+1} \sum_{|j| \leq m} I_n(g(n, \lambda) + \omega_j),$$

where $m = m_n \to \infty$ but $m = o(n)$ as $n \to \infty$ and $g(n, \lambda)$ is the multiple of $2\pi/n$ closest to $\lambda \in [-\pi, \pi]$.

Smoothed periodogram estimator of LRV

- We consider the smoothed periodogram estimator of the LRV given by
$$Q_n = \frac{1}{m}\sum_{j=1}^{m} I_n(\omega_j),$$
where $m = m_n \to \infty$ but $m = o(n)$ as $n \to \infty$.
- We extend the asymptotic normality of Q_n to general nonlinear moving averages, a result previously available only for linear processes.
- Because Q_n relies on a narrow band of local Fourier frequencies, it is relevant for local Whittle estimation of the Hurst exponent.

Weakly dependent nonlinear processes

Linear process

Definition

$\{X_t\}_{t\in\mathbb{Z}}$ is a linear process (or a moving average process) if

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad \text{for each} \quad t \in \mathbb{Z},$$

where

- $\{a_j\}_{j\geq 0} \subset \mathbb{R}$ is such that $\sum_{j=0}^{\infty} a_j^2 < \infty$;
- $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed (iid) random variables with $\mathrm{E}\,\varepsilon_0 = 0$ and $\mathrm{E}\,\varepsilon_0^2 < \infty$.

The asymptotic normality of Q_n under the assumption of linearity follows from the central limit theorem for weighted sums of periodogram ordinates established by Giraitis and Koul (2013).

Nonlinear moving average

Many commonly considered stationary processes are nonlinear (for example, bilinear, threshold, GARCH, stochastic volatility models) and not covered by the currently existing asymptotic results.

Definition

$\{X_t\}_{t\in\mathbb{Z}}$ is a nonlinear moving average if

$$X_t = g(\varepsilon_t, \varepsilon_{t-1}, \dots) \quad \text{for each} \quad t \in \mathbb{Z}, \tag{1}$$

where

- $g : S^\infty \to \mathbb{R}$ is a measurable function;
- $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of iid random elements with values in a measurable space S.

L^p-m-approximable nonlinear moving average

Consider a modified version of the nonlinear moving average in (1) defined by

$$X_t^{(m)} = g(\varepsilon_t, \ldots, \varepsilon_{t-m+1}, \varepsilon'_{t-m}, \varepsilon'_{t-m-1}, \ldots),$$

where $\{\varepsilon'_t\}_{t\in\mathbb{Z}}$ is an independent copy of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$.

Definition

A nonlinear moving average is L^p-m-approximable if

$$\sum_{t=1}^{\infty} \|X_t - X_t^{(t)}\|_p < \infty$$

with $p \geq 1$, where $\|\cdot\|_p = (\mathrm{E}\,|\cdot|^p)^{1/p}$.

Hörmann and Kokoszka (2010)

Limiting distribution of the estimator

Assumption

Assumption 1

$\{X_t\}_{t\in\mathbb{Z}}$ is a nonlinear moving average and

$$\|X_t - X_t^{(m)}\|_4 = O(m^{-\eta}) \quad \text{as} \quad m \to \infty \tag{2}$$

with some $\eta > 3$.

Assumption 1 implies that

- $\{X_t\}_{t\in\mathbb{Z}}$ is L^4-m-approximable;
- $\sum_{h=1}^{\infty} h^r |\gamma(h)| < \infty$ for $0 \le r < \eta - 1$.

Main result

Theorem 1

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a nonlinear moving average such that $\mathrm{E}\, X_0 = 0$ and $\mathrm{E}\, |X_0|^4 < \infty$. Suppose that Assumption 1 is satisfied and $m \to \infty$ but $m = o(n^{4/5})$ as $n \to \infty$. Then

$$\frac{\sqrt{m}}{f(0)}[Q_n - f(0)] \xrightarrow{D} N(0,1) \quad as \quad n \to \infty.$$

Two components of the proof

- The proof is based on the decomposition

$$\frac{\sqrt{m}}{f(0)}[Q_n - f(0)] = \underbrace{\frac{\sqrt{m}}{f(0)}(Q_n - \mathsf{E}\,Q_n)}_{\xrightarrow{D} N(0,1)} + \underbrace{\frac{\sqrt{m}}{f(0)}(\mathsf{E}\,Q_n - f(0))}_{\to 0}.$$

- The key component of the proof of the convergence to $N(0,1)$ in distribution is a general central limit theorem for quadratic forms established by Liu and Wu (2010) (see Theorem 6 therein).

Bias

The rate at which the bias of the estimator Q_n vanishes depends on the rate at which $\gamma(h) \to 0$ as $h \to \infty$.

Theorem 2

Suppose that $\mathrm{E}\, X_0 = 0$, $\mathrm{E}\, X_0^2 < \infty$, $\sum_{h=1}^{\infty} h^r |\gamma(h)| < \infty$ *with some* $r \geq 0$ *and* $m \to \infty$ *but* $m = o(n)$ *as* $n \to \infty$. *Then*

$$\mathrm{E}\, Q_n - f(0) = \begin{cases} o(1) & \text{if } r = 0, \\ O((m/n)^r) & \text{if } 0 < r \leq 1, \\ O(\max\{n^{-1}, (m/n)^{\min\{2,r\}}\}) & \text{if } r > 1. \end{cases}$$

Application to change point analysis

Change-point model

Suppose that $X_1, \ldots, X_n$ are generated as

$$X_t = \begin{cases} \mu + Y_t & \text{if } 1 \leq t \leq k^*, \\ \mu + \Delta + Y_t & \text{if } k^* < t \leq n, \end{cases} \tag{3}$$

where

- k^* is the unknown time at which the potential change point occurs;
- μ is the unknown baseline mean;
- Δ is the magnitude of change at the change point k^*;
- $\{Y_t\}_{t \in \mathbb{Z}}$ is a zero-mean weakly dependent stationary process.

Long-memory model

A stationary process $\{X_t\}_{t\in\mathbb{Z}}$ has long memory if

$$\gamma(h) = h^d L(h) \quad \text{for} \quad h = 1, 2, \ldots, \tag{4}$$

where

- $-1 < d < 0$ controls the decay rate;
- $L : [0, \infty) \to \mathbb{R}$ is a slowly varying function, i.e., an eventually positive or eventually negative measurable function such that

  $$\frac{L(bx)}{L(x)} \to 1$$

  as $x \to \infty$ for each $b > 0$.

Samorodnitsky (2016)

Properties of the long-memory model

- The variance of the first n observations grows faster than linearly

$$\lim_{n \to \infty} \frac{\mathrm{Var}(X_1 + \ldots + X_n)}{n} = \infty.$$

- Even though $\{\gamma(h)\}_{h \in \mathbb{Z}}$ are not absolutely summable, there exists a spectral density that satisfies

$$f(x) \sim \frac{1}{2\Gamma(-d)\cos(\pi d/2)} |x|^{-(1+d)} L(1/|x|) \quad \text{as} \quad x \to 0$$

  with $-1 < d < 0$.
- In particular, the spectral density has a singularity at 0, i.e.,

$$f(x) \to \infty \quad \text{as} \quad x \to 0.$$

Samorodnitsky (2016)

Change-point models versus long-memory models

- How might change-points give rise to patterns resembling long memory?
- We have that
$$\hat{\gamma}(h) \xrightarrow{P} \gamma(h)$$
as $n \to \infty$, but $\gamma(h) \not\to 0$ as $h \to \infty$ for the change-point model provided that $\Delta > 0$ so the correlogram looks like as if the decay rate of the autocovariances is slow.
- Furthermore, $\mathrm{E}\, I(\omega_j)$ typically attains its largest value at $j = 1$ for large values of n creating the impression of a spectral density with a singularity at zero.

Yau and Davis (2012)

- Hence, change-point and long-memory models can produce similar patterns in time plots, correlograms, and periodograms despite being fundamentally different.
- Several approaches have been proposed to distinguish between long-range dependence and change-point models (for example, Berkes et al. (2006), Baek and Pipiras (2012) and Norwood and Killick (2018)).
- We show that the approach of Baek and Pipiras (2012) is also applicable to nonlinear moving averages satisfying the assumptions of Theorem 1.
- The approach of Baek and Pipiras (2012) is based on the local Whittle estimator of the Hurst exponent.

Hurst exponent

Definition

Suppose $\{X_t\}_{t\in\mathbb{Z}}$ is a stationary sequence of random variables and its spectral density satisfies

$$f(\lambda) \sim c|\lambda|^{1-2H} \quad \text{as} \quad \lambda \to 0 \tag{5}$$

with $c > 0$ and $0 < H < 1$. Then H is called the Hurst exponent or the self-similarity parameter of $\{X_t\}_{t\in\mathbb{Z}}$.

- If $H = 1/2$, $\{X_t\}_{t\in\mathbb{Z}}$ has short memory.
- if $H > 1/2$, $\{X_t\}_{t\in\mathbb{Z}}$ has long memory.
- The long-memory model given in (4) without a slowly varying function satisfies (5) with $H = 1 + d/2 \in (1/2, 1)$.

Local Whittle estimator of the Hurst exponent

Definition

The local Whittle estimator (LWE) of the Hurst exponent H is defined as

$$\hat{H} = \underset{H \in \Theta}{\arg\min}\, L(H)$$

where $\Theta = [\Delta_1, \Delta_2]$ with $0 < \Delta_1 < \Delta_2 < 1$,

$$L(H) = \log\left\{\frac{1}{m}\sum_{l=1}^{m} \omega_l^{2H-1} I_n(\omega_l)\right\} - (2H-1)\frac{1}{m}\sum_{l=1}^{m} \log \omega_l,$$

and m is the number of low frequencies used in the estimation.

Δ_1 and Δ_2 can be taken arbitrarily near 0 and 1, or set to reflect prior beliefs (e.g., $\Delta_1 = 1/2$).

Robinson (1995)

Hypotheses

The objective is to test

$$H_0 : \text{the observations } \{X_1, \ldots, X_n\} \text{ follow the change-point model}$$

versus

$$H_1 : \text{the observations } \{X_1, \ldots, X_n\} \text{ follow the long-memory model.}$$

A potential change point can severely bias the LWE so its effect is removed by considering the residuals

$$\hat{R}_t = \begin{cases} X_t - \hat{k}^{-1} \sum_{i=1}^{\hat{k}} X_i, & 1 \le t \le \hat{k}, \\ X_t - (n - \hat{k})^{-1} \sum_{i=\hat{k}+1}^{n} X_i, & \hat{k} + 1 \le t \le n, \end{cases}$$

where $\hat{k}$ is the standard CUSUM estimator

$$\hat{k} = \min\left\{ k : \left| \sum_{j=1}^{k} X_j - \frac{k}{n} \sum_{j=1}^{n} X_j \right| = \max_{1 \le s \le n} \left| \sum_{j=1}^{s} X_j - \frac{s}{n} \sum_{j=1}^{n} X_j \right| \right\}.$$

Test statistic

The test statistic is

$$T^{(\hat{R})} = 2\sqrt{m}\Big(\hat{H}^{(\hat{R})} - \frac{1}{2}\Big),$$

where $\hat{H}^{(\hat{R})}$ is the LWE based on the residuals $\{\hat{R}_t\}_{1\leq t\leq n}$.

Distribution under H_0

Theorem 3

Suppose that model (3) holds with $\{Y_t\}_{1\le t\le n}$ satisfying $\operatorname{E} Y_1 = 0$ as well as $\operatorname{E}|Y_1|^4 < \infty$ and Assumption 1. In addition, assume that

(a) $k^* = \lfloor \theta n \rfloor$ *with some* $\theta \in (0,1)$*;*

(b) *the change in mean level* $\Delta = \Delta(n)$ *and the change point estimator* $\hat{k}$ *satisfy*

$$n\Delta^2 \to \infty, \quad \Delta^2|\hat{k} - k^*| = O_p(1) \quad \text{as} \quad n \to \infty;$$

(c)

$$\frac{1}{m} + \frac{m^5(\log m)^2}{n^4} + \frac{m(\log m)^2}{n\Delta^2} \to 0 \quad \text{as} \quad n \to \infty.$$

Then $T^{(\hat{R})} \xrightarrow{D} N(0,1)$ *as* $n \to \infty$.

Idea of the proof of Theorem 3

The proof of Theorem 3 can be reduced to known arguments, Theorem 1 and the following convergence that we establish for nonlinear moving averages and under certain additional assumptions

$$\frac{1}{\sqrt{m}} \sum_{j=1} \nu_{l,m} \Big(\frac{I_n(\omega_j)}{f(0)} - 1 \Big) \xrightarrow{D} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty,$$

where

$$\nu_{l,m} = \log l - \frac{1}{m} \sum_{l'=1}^{m} \log l'$$

are the specific weigths that follow from the LWE.

Rejection rule

Provided that certain assumptions hold,

$$T^{(\hat{R})} \xrightarrow{D} N(0,1) \quad \text{as} \quad n \to \infty$$

under H_0 and hence we reject H_0 of the change-point model if

$$T^{(\hat{R})} > z_{1-\alpha},$$

where $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of $N(0,1)$.

GARCH

Definition

$\{Y_t\}_{t\in\mathbb{Z}}$ is said to be a GARCH$(1,1)$ process if it satisfies the equations

$$Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where

- $\alpha_0 > 0$, $\alpha_1, \beta_1 \geq 0$, $\alpha_1 + \beta_1 < 1$;
- $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are iid random variables such that $\mathrm{E}\,\varepsilon_0 = 0$, $\mathrm{E}\,\varepsilon_0^2 = 1$, and $\mathrm{E}\,\varepsilon_0^4 < \infty$.

- $\alpha_1 + \beta_1 < 1 \Rightarrow$ the GARCH$(1,1)$ process has a stationary solution.
- $\mathrm{E}\,Y_0^4 \iff \mathrm{E}\,|\beta_1 + \alpha_1 \epsilon_0^2|^2 < 1$
- $\mathrm{E}\,|\beta_1 + \alpha_1 \epsilon_0^2|^2 < 1 \implies$ Assumption 1.

Empirical size

Nominal size	1%			5%			10%		
n	1000	5000	10000	1000	5000	10000	1000	5000	10000
				$\Delta = 0.30, \theta = 0.25$					
$m = n^{0.60}$	0.90	1.12	1.68	4.00	5.00	6.48	7.86	9.96	11.56
$m = n^{0.70}$	0.92	1.18	1.46	3.98	5.00	5.58	7.86	9.68	9.90
				$\Delta = 0.30, \theta = 0.50$					
$m = n^{0.60}$	0.64	0.68	0.78	3.14	3.36	4.20	6.26	6.58	8.28
$m = n^{0.70}$	0.80	0.70	0.98	3.38	3.96	4.06	6.68	7.36	7.66
				$\Delta = 0.30, \theta = 0.75$					
$m = n^{0.60}$	0.76	1.08	1.72	3.66	5.02	5.96	7.26	10.00	11.08
$m = n^{0.70}$	0.80	1.12	1.62	3.88	4.90	5.66	7.80	9.38	10.18
				$\Delta = 0.60, \theta = 0.25$					
$m = n^{0.60}$	1.68	1.76	1.58	6.24	6.04	6.00	11.12	11.10	11.32
$m = n^{0.70}$	1.52	1.44	1.64	5.70	6.06	6.04	10.78	11.04	10.24
				$\Delta = 0.60, \theta = 0.5$					
$m = n^{0.60}$	0.78	0.74	0.72	3.82	3.74	4.32	7.68	7.24	8.62
$m = n^{0.70}$	0.86	0.84	1.00	3.98	4.18	4.20	7.60	7.82	8.12
				$\Delta = 0.60, \theta = 0.75$					
$m = n^{0.60}$	1.58	1.82	1.66	6.16	6.32	6.26	10.84	11.14	11.46
$m = n^{0.70}$	1.32	1.58	1.94	5.86	6.12	5.98	10.86	10.48	10.66

Summary

Summary

- We consider the smoothed periodogram estimator of the LRV.
- The limiting distribution of the estimator is derived for nonlinear moving averages, a result previously available only for linear processes.
- We demonstrate how our result can be used in local Whittle estimation of the Hurst exponent and in distinguishing change-point models from long-memory models.
- For more details, see Characiejus, Kokoszka, and Meng (2025).

`https://imada.sdu.dk/u/characiejus/`

References (1/2)

Baek, Changryong and Vladas Pipiras (Jan. 2012). **"Statistical tests for a single change in mean against long-range dependence".** In: *Journal of Time Series Analysis* 33.1, pp. 131–151.

Berkes, István et al. (2006). **"On discriminating between long-range dependence and changes in mean".** In: *The Annals of Statistics* 34.3, pp. 1140–1165.

Characiejus, Vaidotas, Piotr Kokoszka, and Xiangdong Meng (2025). **"Estimation of the Long-Run Variance of Nonlinear Time Series With an Application to Change Point Analysis".** In: *Journal of Time Series Analysis.*

Giraitis, Liudas and Hira L. Koul (2013). **"On asymptotic distributions of weighted sums of periodograms".** In: *Bernoulli* 19.5B, pp. 2389–2413.

Hörmann, Siegfried and Piotr Kokoszka (2010). **"Weakly dependent functional data".** In: *The Annals of Statistics* 38.3, pp. 1845–1884.

Liu, Weidong and Wei Biao Wu (2010). **"ASYMPTOTICS OF SPECTRAL DENSITY ESTIMATES".** In: *Econometric Theory* 26.4, pp. 1218–1245.

Norwood, Ben and Rebecca Killick (2018). **"Long memory and changepoint models: a spectral classification procedure".** In: *Statistics and Computing* 28.2, pp. 291–302.

References (2/2)

Robinson, P. M. (1995). **"Gaussian Semiparametric Estimation of Long Range Dependence".** In: *The Annals of Statistics* 23.5, pp. 1630–1661.

Samorodnitsky, Gennady (2016). ***Stochastic Processes and Long Range Dependence.*** Springer Series in Operations Research and Financial Engineering. Springer Cham.

Yau, Chun Yip and Richard A. Davis (July 2012). **"Likelihood inference for discriminating between long-memory and change-point models".** In: *Journal of Time Series Analysis* 33.4, pp. 649–664.