IDEI Working Paper Series n° 633

September, 2009

"Nonparametric Beta Kernel Estimator for Long Memory Time Series"

Taoufik BOUEZMARNI and Sébastien VAN BELLEGEM

Institut d'Economie Industrielle (IDEI) – Manufacture des Tabacs
Aile Jean-Jacques Laffont – 21, allée de Brienne – 31000 TOULOUSE – FRANCE
Tél. + 33(0)5 61 12 85 89 – Fax + 33(0)5 61 12 86 37 – **www.idei.fr** – contact@cict.fr

INSTITUT D'ECONOMIE INDUSTRIELLE

Nonparametric Beta Kernel Estimator for Long Memory Time Series*

Taoufik Bouezmarni — McGill University
Sébastien Van Bellegem[1] — Toulouse School of Economics

This version: September, 11, 2009

Abstract

The paper introduces a new nonparametric estimator of the spectral density that is given in smoothing the periodogram by the probability density of Beta random variable (Beta kernel). The estimator is proved to be bounded for short memory data, and diverges at the origin for long memory data. The convergence in probability of the relative error and Monte Carlo simulations suggest that the estimator automaticaly adapts to the long- or the short-range dependency of the process. A cross-validation procedure is also studied in order to select the nuisance parameter of the estimator. Illustrations on historical as well as most recent returns and absolute returns of the S&P500 index show the reasonable performance of the estimation, and show that the data-driven estimator is a valuable tool for the detection of long-memory as well as hidden periodicities in stock returns.

Keywords: Spectral density, Long range dependence, Nonparametric estimation, Periodogram, Kernel smoothing, Beta kernel, Cross-validation

1 Introduction

The estimation of a spectral density often requires to know whether the observed stationary time series is short or long memory. Long memory, or long range dependent time series is characterized by a spectral density that is unbounded at frequency zero, therefore the choice of an optimal nonparametric estimator will be different if the spectral density is bounded or not. It is one goal of the present paper to go beyond that limitation and to propose an estimator that is applicable to any stationary data, being long range dependent or not.

*This work was supported by the "Agence National de le Recherche" under contract ANR-09-JCJC-0124-01.

[1]Corresponding author. Email address: svb@tse-fr.eu

A well-established nonparametric estimation procedure consists in estimating first the parameter d_0 of the long memory process. In that approach, the spectral density f is assumed to behave like

$$f(\lambda) = |\lambda|^{-2d_0} L(\lambda)$$

as $\lambda \to 0+$, for $d_0 \in (-1/2, 1/2)$, where $L(\lambda)$ is slowly varying and such that $0 < L(0) < \infty$. Many papers study the estimation of d_0. Among recent advances we can cite the approaches of Andrews and Sun (2004), Robinson and Henry (2003) or Henry (2007) to name but a few. See also the recent surveys in Doukhan, Oppenheim, and Taqqu (2003); Robinson (2003); Palma (2007).

Inference on d_0 allows to test whether that parameter is significantly larger than zero, that is if the process is long memory, see Lobato and Robinson (1998), Lobato and Velasco (2000) or Ohanissian, Russel, and Tsay (2008). The testing step is important because the asymptotic distribution of the spectral density estimator is usually not the same if $d_0 = 0$ or if $d_0 > 0$. If the process is short memory, the nonparametric estimation of its spectral density becomes a classical problem of inference. If not, it has been proposed to estimate the spectrum for λ close to zero by $\hat{C}|\lambda|^{-2\hat{d}_0}$ for a consistent estimator of $\hat{d}_0$ and where $\hat{C}$ is another estimator that makes the overall estimation consistent (the procedure is recalled with more details in Section 3.3 below). Away from the origin, another nonparametric estimation must be used in order to evaluate the spectrum for $\lambda > 0$.

In this paper we study a new nonparametric estimator of the spectral density that is given by a smoothing of the periodogram by a Beta kernel. The Beta kernel is the probability density function of a Beta random variable. It is not a symmetric kernel, and its shape varies according tot the frequency where the spectrum is estimated, see Section 2 below. Beta kernel smoothing was introduced by Brown and Chen (1999) in the context of smoothing the Bernstein polynomials in order to estimate compactly supported regression curves. It has then been used in order to address the boundary bias problem in the context of regression or probability density estimation, see Chen (1999); Chen (2000).

Because the Beta kernel diverges at zero when its bandwidth shrinks, it is an appealing smoother of the periodogram when the process is long memory. In fact, we show below that it adapts automatically to the memory of the time series: If the process is short memory, the resulting estimation of the spectral density is automatically bounded, whereas the estimator diverges at the origin when it is applied to long range dependent data.

The paper is organized as follows. In Section 2 we define the Beta kernel estimator of the spectral density, and provide an illustration in the estimation of the returns and absolute returns of the S&P500 index. The properties of the estimator are discussed in Section 3. First, we study its behavior outside the origin and establish its uniform convergence over any compact set of frequencies. Then we consider what happens at the origin, and show that the estimator is bounded in probability at the origin for long range data, and unbounded for short memory processes. We also derive a stronger result that is the relative convergence of the estimator at the origin. Next, we study the finite sample performance of the estimator. A Monte Carlo study on three parametric (ARFIMA) models confirms the reasonable adaptation of the proposed estimator to the rande of memory of the process. We compare the empirical performances of our estimator with the semi-nonparametric estimator of Robinson (1995) and show the merits of both methods.

The last Section addresses more practical aspects of the estimation procedure. As every nonparametric estimators, the Beta kernel smoother depends on a nuisance parameter. In Section 4 we study a cross-validation method to select that parameter following the general method of Hurvich (1980). Another Monte Carlo study demonstrates the good performance of the fully data-driven Beta kernel estimator, which is also illustrated on more recent paths of the S&P500 index. An appendix contains the proofs of all results.

2 The Beta kernel estimator of the spectral density

2.1 Construction of the estimator

Beta kernel estimators were studied by Chen (2000) in the context of the estimation of regression curves. The motivation was to develop a kernel smoothing technique that is free of boundary bias. In the context of time series analysis, this property is valuable since the nonparametric kernel estimator of the periodogram is not necessarily adapted at the border, especially if there is a pole at frequency $\lambda = 0$.

We first construct the estimator. Suppose we observe $X_1, \ldots, X_T$ from a stationary process with spectral density $f(\lambda) = \sum_k \gamma(k) \exp(-2\pi i \lambda k)$ where $\gamma(k)$ is the covariance function of X_t. For the sake of simplicity, we assume the stationary process to be zero mean. The periodogram

$$I_T(\omega_j) = \frac{1}{2\pi T} |\sum_{t=1}^{T} X_t \exp(-2\pi i \omega_j t)|^2 \qquad \omega_j = \frac{j}{T}, j = 1, \ldots, T/2.$$

is known to be an asymptotically unbiased, not consistent estimator of the spectral density f. A consistent estimator is found after an appropriate smoothing of I_T over frequencies. In this paper, we study the estimator

$$\hat{f}(\lambda) = \frac{1}{T} \sum_{j=1}^{T} K_{b,\lambda}(\omega_j) I_T(\omega_j) \tag{2.1}$$

where $K_{b,\lambda}$ is a *Beta-kernel* defined as

$$K_{b,\lambda}(\omega) = \frac{\omega^{\lambda/b}(1-\omega)^{(1-\lambda)/b}}{B\left(\frac{\lambda}{b}+1, \frac{1-\lambda}{b}+1\right)} \mathbb{1}_{0 \leqslant \omega \leqslant 1}$$

for the beta function B and the smoothing parameter b. The Beta-kernel is the probability density function of a Beta$\{1 + \lambda/b, 1 + (1-\lambda)/b\}$ random variable.

In contrast to most kernel estimators, the estimator $\hat{f}(\lambda)$ does not use a symmetric kernel but a kernel whose shape varies with λ. That property is illustrated at Figure 1, where the function $K_{b,\lambda}(\cdot)$ is displayed for some frequencies λ. This varying shape kernel implies that the amount of smoothing changes according to the frequency where spectrum is estimated. As noticed by Chen (1999), the variance of the Beta$\{1 + \lambda/b, 1 + (1-\lambda)/b\}$ random variable is of order

$$b\lambda(1-\lambda) + O(b^2)$$

suggesting that the amount of smoothing is small at the border of the support. Note also that the beta kernel does not put any weight outside the support of $f(\lambda)$.

2.2 Empirical illustration

An eminent feature of the Beta kernel estimator is its adaptivity to the boundness or unboundness of the spectrum at the origin $\lambda = 0$. To illustrate that property, we consider in Figure 2 a segment of the daily absolute returns of the S&P500 that was analysed by Lobato and Savin (1998). Using a Lagrange multiplier test, the later conclude that there is no evidence of long memory in the levels of the returns, whereas their analysis favors long memory of the squared returns.

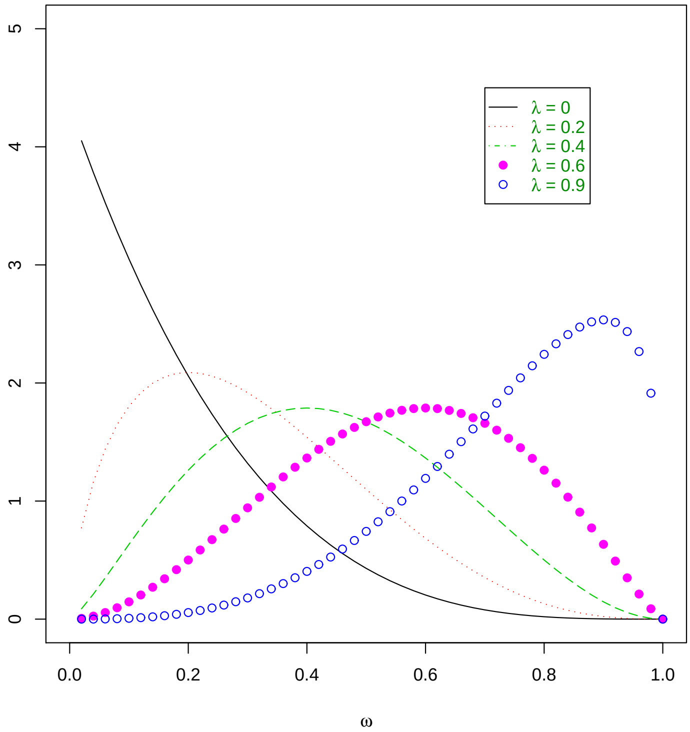


Figure 1: Beta kernel $K_{b,\lambda}(\cdot)$ used to estimate the spectral density $f(\lambda)$. The shape of that kernel varies according to the frequency λ where the spectral density is estimated ($b = 0.3$).

In Figure 3(a) and (b), we display the empirical autocorrelation function of the log returns, and absolute log returns respectively. Those pictures illustrate the conclusions of Lobato and Savin (1998) recalled above.

Estimation of the log-spectrum by Beta kernel of the log returns and the absolute log returns is proposed in Figures 3(c) and (d) respectively. The estimator is drawn for several values of the smoothing parameters, $b = 0.005, 0.01$ and 0.05. We observe that smaller b is, more oscillating is the estimator. We therefore recover the usual regularity properties of the estimator with respect to b. A data-driven choice of b is proposed in Section 4 below.

Figures 3(c) and (d) also show that the Beta kernel estimator of the spectrum is bounded for the log returns, and is diverging for the absolute log returns. This illustrates how the estimator automatically adapts to the unknown memory structure of the process. In other words, the estimator can be applied to time series of any type of memory, in contrast to most estimators who are applicable either to short or to long memory processes.

For the sake of comparison, other kernel smoothing of the periodogram are displayed in Figures 3(e) and (f), respectively for the log-returns and the absolute log returns. Three

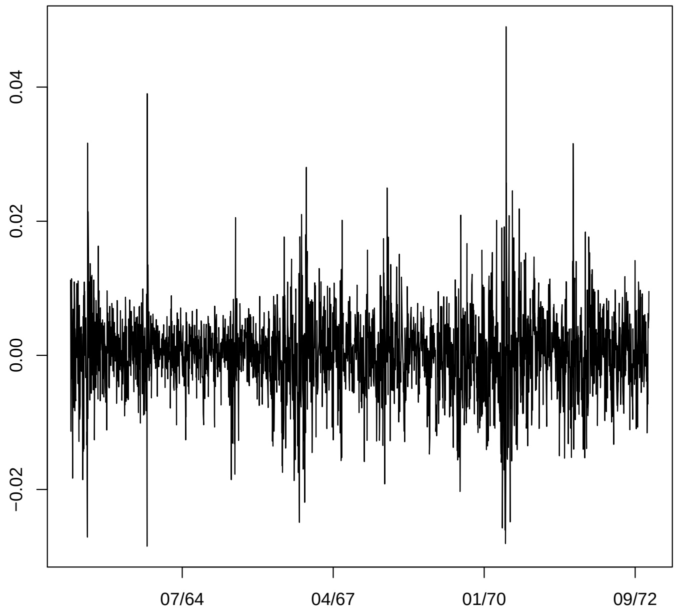
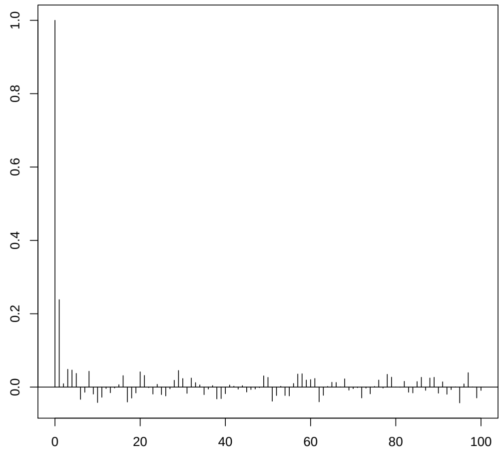
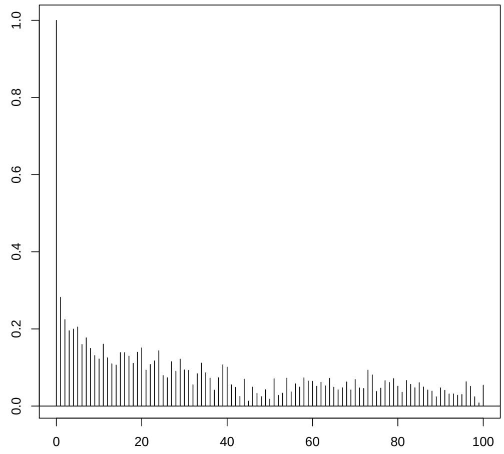


Figure 2: Daily log returns of S&P500, July 1962 to December 1972, 2616 data

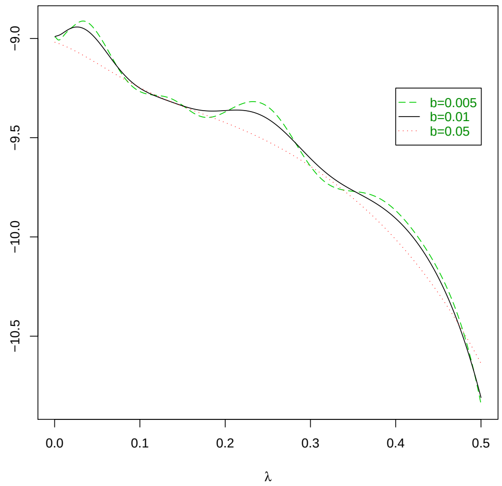
kernel smoothing are superimposed: (i) The symmetric Daniell kernel estimator with bandwidth 0.036; (ii) a rectangular kernel estimator with bandwidth 0.043 and (iii) an asymmetric triangular kernel. For exact definitions, we refer e.g. to Brillinger (2001) or many other textbooks. Although the standard kernel methods show a peak close to the frequency zero in Figure 3(f), it is appearent that the unboundness of the spectrum is more difficult to display with classical methods. We could vary the bandwidth in order to underline the peak close to the pole, but, in such a case, the quality of estimation far from frequency zero would be very weak.



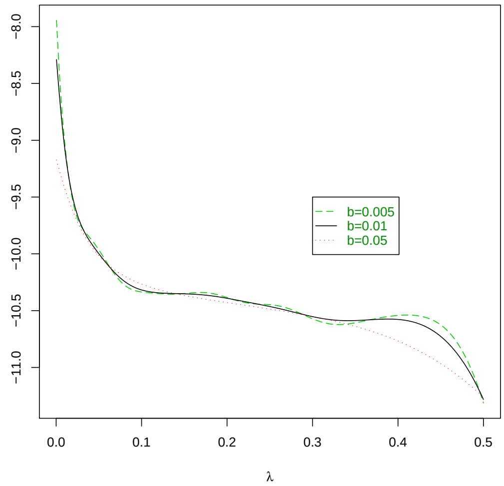
(a) Empirical autocorrelation function of the log returns



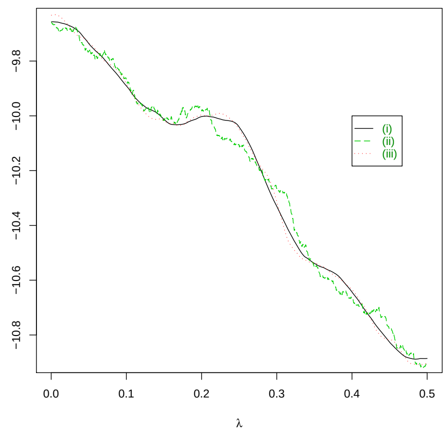
(b) Empirical autocorrelation function of the absolute log returns



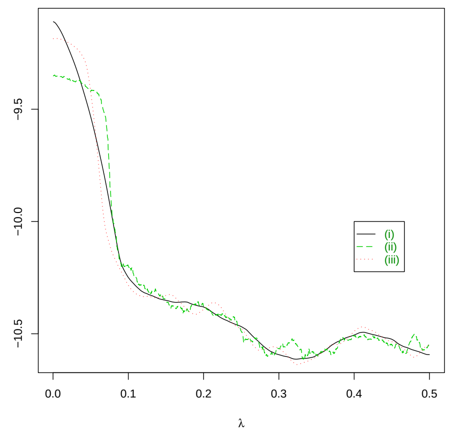
(c) Log spectrum of the log returns, estimated by Beta kernel estimator with various bandwidths b



(d) Log spectrum of the absolute log returns, estimated by Beta kernel estimator with various bandwidths b



(e) Log spectrum of the log returns, estimated by other kernel smoothing of the periodogram: (i) Symmetric Daniell kernel; (ii) Rectangular kernel; (iii) Asymmetric triangular kernel.



(f) Log spectrum of the absolute log returns, estimated by other kernel smoothing of the periodogram as in (e) (details in text)

Figure 3: Empirical autocorrelation function and Log spectrum estimation of the daily log returns of S&P500 (July 1962 to December 1972)

3 Properties of the estimator

In this section, we explore the asymptotic and the finite sample properties of the estimator. Overall, we assume that the process can be long memory, in the sense that it may have a pole at the origin:

Assumption 3.1. *The spectral density f is such that*[1] $f(\lambda) \sim \lambda^{-\beta} g(\lambda)$ *as* $\lambda \to 0+$, *for* $0 \leqslant \beta < 1$, *where g is a Lipschitz, continuous, bounded, strictly positive function on* $[0,1]$.

We start our discussion by studying the behavior of the estimator outside the origin.

3.1 Behavior of the estimator outside the origin

Given a stationary, zero mean time series $\{X_t; t = 1, \ldots T\}$ with a spectral density $f(\lambda)$ that is two times differentiable, we derive below the appropriate rate of convergence of the bandwidth $b = b(T)$ such that the bias and the variance of the Beta kernel estimator vanish asymptotically. We also prove the uniform convergence of the estimator on any compact set in $(0, 1)$.

The expectation of the Beta kernel estimator at frequency $\lambda \neq 0$ is given by

$$\mathbb{E}(\hat{f}(\lambda)) = \int_0^1 K_{b,\lambda}(u) f(u) du + R_1 + R_2$$

where R_1 and R_2 are two approximation terms that are given by

$$R_1 = \frac{1}{T} \sum_j K_{b,\lambda}(\omega_j) f(\omega_j) - \int_0^1 K_{b,\lambda}(u) f(u) du$$

and

$$R_2 = \mathbb{E}\Big[\frac{1}{T} \sum_j K_{b,\lambda}(\omega_j)(I_T(\omega_j) - f(\omega_j))\Big] .$$

By the smoothness assumption on f, the approximation term R_1 has rate $O(T^{-1})$. The term R_2 pools the periodogram over frequencies, and can be computed e.g. using the results of Robinson (1994b). Noting that $K_{b,\lambda}(u) \leqslant c_1 b^{-1/2} \{\lambda(1-\lambda)\}^{-1/2}$ where c_1 is a positive constant (see Chen (2000)), Proposition 2 in Robinson (1994b) leads to

$$R_2 = O\left(\frac{1}{\sqrt{b}} \left(\frac{m_T}{T}\right)^{1-\beta}\right)$$

[1] For two functions $h_1(\lambda)$ and $h_2(\lambda)$, we write $h_1(\lambda) \sim h_2(\lambda)$ if there exists two nonegative, finite constant c_1 and c_2 such that $c_1 \leqslant h_1(\lambda)/h_2(\lambda) \leqslant c_2$ for all λ.

with $m_T^{-1} + m_T/T \to 0$. If the sequence m_T is selected such that $R_2 = o(1)$ we can write

$$\begin{aligned}\mathbb{E}(\hat{f}(\lambda)) &= \int_0^1 K_{b,\lambda}(u) f(u) du + o(1) \\ &= \mathbb{E}(f(\xi_\lambda)) + o(1)\end{aligned} \tag{3.1}$$

where ξ_λ is a Beta$\{1 + \lambda/b, 1 + (1-\lambda)/b\}$ random variable. Under standard smoothness conditions on f, Chen (2000) has derived the general approximation bias of $\mathbb{E}(f(\xi_\lambda))$, leading to

$$f(\lambda) - \mathbb{E}(\hat{f}(\lambda)) = b\{(1-2\lambda)f'(\lambda) + \frac{1}{2}\lambda(1-\lambda)f''(\lambda)\} + o(b) + o(1).$$

We proceed similarly for the variance of estimator. From Robinson (1995) and Moulines and Soulier (1999), under technical conditions the periodogram of long memory time series is such that $\mathbb{C}\mathrm{ov}(I_T(\omega_s)/f(\omega_s), I_T(\omega_t)/f(\omega_t)) = r_{st}$ where $\sum_{s<t} r_{st} = O(\log^r(T))$ for some $r > 0$. After appropriate approximations that are analogous to the above calculation of the bias, we find

$$\mathbb{V}\mathrm{ar}(\hat{f}(\lambda)) = A(\lambda)\mathbb{E}(f(\rho_\lambda)^2) + O\left(\frac{1}{b}\frac{\log^r(T)}{T^{2-2\beta}}\right)$$

where ρ_λ is a Beta$\{1 + 2\lambda/b, 1 + 2(1-\lambda)/b\}$ random variable and

$$A(\lambda) = \frac{B\left(\frac{2\lambda}{b} + 1, \frac{2(1-\lambda)}{b} + 1\right)}{B\left(\frac{\lambda}{b} + 1, \frac{1-\lambda}{b} + 1\right)^2}.$$

If $\beta < 1/2$ and with an appropriate bandwidth b (that is written below), the reminder term is negligible and converges to zero. From a Taylor expansion and the asymptotic properties of $A(\lambda)$ (cf Chen (2000)) the variance of the estimator is found to be

$$\mathbb{V}\mathrm{ar}(\hat{f}(\lambda)) = \begin{cases} \frac{1}{Tb^{1/2}} \frac{f(\lambda)^2}{2\sqrt{\pi\lambda(1-\lambda)}} + o((Tb^{1/2})^{-1}) + o(1) & \text{if} \quad \lambda/b \text{ and } (1-\lambda)/b \to \infty \\ \frac{C(\kappa)}{Tb}\{f(\lambda)^2 + O(T^{-1})\} + o(1) & \text{if} \quad \lambda/b \text{ or } (1-\lambda)/b \to \kappa. \end{cases} \tag{3.2}$$

where κ is a strictly positive constant and

$$C(\kappa) = \frac{\Gamma(2\kappa + 1)}{2^{1+2\kappa}\Gamma(\kappa+1)^2}.$$

Considering the bias and the variance convergences, we check that outside the origin the Beta kernel estimator is asymptotically unbiased estimator with vanishing variance if the bandwidth is such that

$$b + \frac{1}{T^{1-2\beta}\sqrt{b}} \to 0 \tag{3.3}$$

if $\beta < 1/2$. The last constraint on β imposes that the spectrum is still square integrable around the pole, and therefore the mean square error is invariant to the explicit variation of f around frequency 0. Following Robinson (1994a), it is possible to go beyond that constraint under more assumptions on f, but leading to different expressions for the mean square error.

In the next result, we state the uniform convergence of the estimator on a compact set outside the origin.

Proposition 3.1. *Let f be the spectral density function such that Assumption 3.1 is fulfilled with $\beta \in [0, 1/2)$. For any compact set I in $(0, 1)$ and if the bandwidth satisfies (3.3) and is such that $(b^{2+\epsilon}T) \to \infty$ for some $\epsilon > 0$, then the beta kernel spectral estimator is uniformly convergent over I, i.e.*

$$\sup_{\lambda \in I} \left| \hat{f}(\lambda) - f(\lambda) \right| \xrightarrow{P} 0.$$

The proof of the proposition is to be found in the Appendix. Note that the result is also valid in the particular case where $\beta = 0$, that is the process is short memory.

3.2 Behavior of the estimator close to the pole

One special interest is to study the behavior of the estimator close to the zero frequency, where the spectrum is not bounded. The first result shows that the Beta kernel estimator for long memory time series is unbounded at the origin.

Proposition 3.2. *Let f be the spectral density function such that Assumption 3.1 is fulfilled with $\beta \in (0, 1/2)$ and consider the Beta kernel estimator (2.1) with a bandwidth that satisfies (3.3) and the two following constraints (i) $Tb^{1+2\beta} \to \infty$ and (ii) $T^{1-2\beta}b \to \infty$. Then the estimator is such that $\hat{f}(0) \xrightarrow{P} +\infty$ as $T \to \infty$.*

The next corollary states the consistency of the estimator at $\lambda = 0$ for short time series.

Corollary 3.1. *Let f be the spectral density function such that Assumption 3.1 is fulfilled with $\beta = 0$ (short memory process). If b satisfies (3.3) and $Tb \to \infty$, then the Beta kernel spectral estimator (2.1) is such that $\hat{f}(0) \xrightarrow{P} f(0)$.*

We conclude that the Beta kernel estimator is automatically adapted to the "type of memory" of the spectral density (long vs short range). This result has been already illustrated in Figure 3.

However, even if the estimator is consistent at the pole, the last proposition does not give any information about the closeness of the estimator to the true value close to the origin. In order to have an idea about that closeness, the next proposition tells more about the *relative* convergence of the Beta kernel estimator when the spectral density is estimated near the origin. In the next section, we also show empirically the reasonable relative rate of convergence of the estimator close to the pole.

PROPOSITION 3.3. *Let f be the spectral density function such that Assumption 3.1 is fulfilled with $\beta \in [0, 1/2)$ and consider the Beta kernel estimator (2.1) with a bandwidth that satisfies (3.3) and $Tb \to \infty$. Then the Beta kernel spectral estimator (2.1) is such that*

$$\left| \frac{\hat{f}(\lambda)}{f(\lambda)} - 1 \right| \xrightarrow{P} 0$$

when λ tends to zero such that $\lambda/b \to \kappa > 0$.

3.3 Finite sample properties

In this section, we examine the properties of the estimator through Monte Carlo simulations.

In order to judge quality of the estimator, we provide a comparison with the semi-parametric estimator of Robinson (1994b). That approach assumes that the spectrum is such that $f(\lambda) \sim C_f \lambda^{1-2H}$ as $\lambda \to 0+$, and proposes consistent estimates of H and C_f that we recall now. Observing that the spectral distribution, $F(\lambda) = \int_0^\lambda f$, is such that $F(q\lambda)/\lambda \sim q^{2(1-H)}$ for all $q \in (0, 1)$, Robinson (1994b) has suggested to estimate H by

$$\hat{H} = 1 - \frac{\log(\hat{F}(q\lambda_m)/\lambda_m)}{2 \log q}$$

for a given q and frequency $\lambda_m = m/T$. Similarly, observing that the spectral distribution is $F(\lambda) = \frac{C_f}{2-2H} \lambda^{2-2H}$, an estimate of C_f is given by

$$\hat{C}_f = 2(1 - \hat{H}) \hat{F}(\lambda_m) \lambda_m^{2(\hat{H}-1)}.$$

Finally, an estimator of the spectrum close to the origin is given by $\hat{f}(\lambda) = \hat{C}_f \lambda^{1-2\hat{H}}$.

The semiparametric estimator depends on the choice of two parameters, q and m. In our computations below, we set $q = 0.5$ as it is often observed in the literature. The choice of m is however more delicate. Based on the expansion of the asymptotic mean square error, some rules for the choice of m have been proposed in Robinson (1994a). They were the starting point of the feasible, data-driven proposal of Delgado and Robinson (1996a) and Delgado and Robinson (1996b).

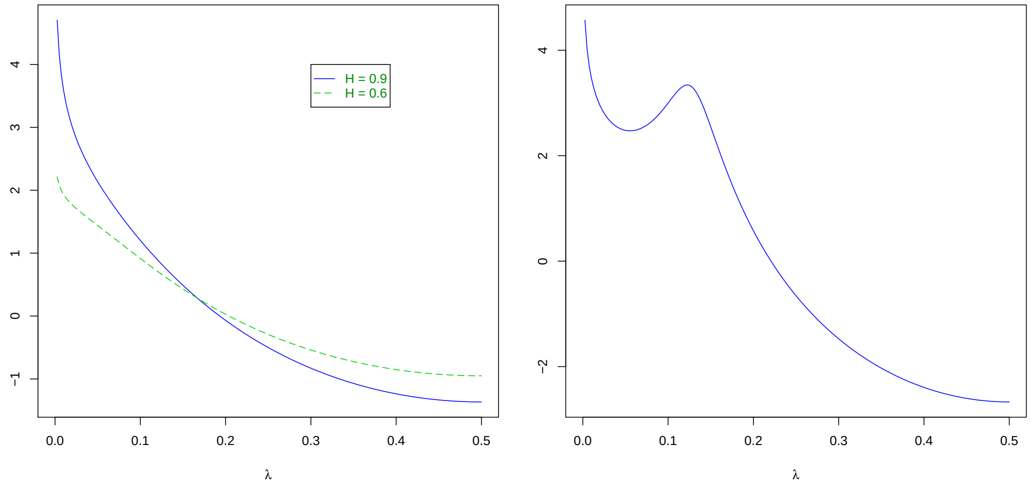
In order to facilitate the comparison between the semiparametric estimator and the Beta kernel estimator, we have simulated below three ARFIMA models that are also studied in Delgado and Robinson (1996a). ARFIMA models provide a well-established parametric specification of long memory. It is given by the fractional autoregressive integrated moving average FARIMA (p, d, q) model that has spectral density

$$f(\lambda) = |1 - \exp(i\lambda)|^{1-2H} h(\lambda), \qquad -1 \leqslant \lambda \leqslant 1,\ H \in [0, 1] \tag{3.4}$$

where

$$h(\lambda) = \sigma^2 \frac{|b(\exp(i\lambda))|^2}{|a(\exp(i\lambda))|^2}$$

with $a(z) = 1 - \sum_{j=1}^{p} a_j z^j$ and $b(z) = 1 - \sum_{j=1}^{q} b_j z^j$. In that model, $H = 1/2$ corresponds to short memory if we assume $0 < h(\lambda) < \infty$, wheras $H > 1/2$ leads to a long memory process. Figure 4 displays the logarithm of the spectral density of the three ARFIMA generating models used in the simulation below.



(a) ARFIMA(1,H,0) with $a_1 = 0.5$ and two values of the memory parameter H.

(b) ARFIMA(2,$H = 0.9$,0) with $a_1 = 1.172$, $a_2 = -0.707$.

Figure 4: Logarithm of the three spectral densities of the ARFIMA used in the simulations.

The estimators were computed on 1000 Monte Carlo simulations of the models, for sample sizes $T = 400, 600, 1000$. Since the semiparametric estimator is a local estimator around the pole, we do not compare with the Beta kernel estimator over all frequencies but only in a neighbourhood of the frequency zero . According to the theory of Robinson (1994b), we compute the error of estimation on the frequencies in $(\lambda_{j_0}, \lambda_{j_1})$, where $j_0 = [\sqrt[5]{T}]$

and $j_1 = [\sqrt{T}]$. The empirical error that we compute is the relative mean absolute deviation, i.e.

$$\text{RMAD}_{[j_0,j_1]} = \frac{1}{j_1 - j_0 + 1} \sum_{j=j_0}^{j_1} \frac{|\hat{f}\left(\frac{j}{T}\right) - f\left(\frac{j}{T}\right)|}{f\left(\frac{j}{T}\right)}$$

where $\hat{f}$ denotes the considered estimator of the spectrum. Taking the relative MAD instread of the MAD is motivated by the unboundesss of the spectral density at frequency zero.

Note that, because j_0 and j_1 depend on T, the range of frequencies where the error is computed is different for each sample size. Therefore the MAD presented in the empiricial study below are only comparable for a given sample size.

Tables 1 to 3 display the results of the Monte Carlo study. The ARFIMA time series were generated via the library 'fracdiff' in R. In order to avoid the dependence of our conclusions to the choice of the bandwidths, we have computed the RMAD for a range of bandwidths. The range of bandwidths b of the Beta kernel estimator is $[0.01, 0.5]$, and range of m in the semiparametric estimation is $[T^{1/2}, T^{4/5}]$. The tables display the five results that were the closest to the best RMAD found. (In the next section, we also address the question of the data driven choice of b.)

	$T = 400$			$T = 600$			$T = 1000$		
Beta kernel:									
	b	RMAD		b	RMAD		b	RMAD	
	0.08	0.416	(0.190)	0.08	0.364	(0.144)	0.08	0.324	(0.104)
	0.115	0.339	(0.130)	0.115	0.302	(0.091)	0.115	0.288	(0.059)
	0.15	0.307	(0.089)	0.15	0.287	(0.061)	0.15	0.298	(0.052)
	0.185	0.299	(0.066)	0.185	0.296	(0.055)	0.185	0.329	(0.065)
	0.22	0.304	(0.060)	0.22	0.317	(0.063)	0.22	0.366	(0.077)
Semiparametric:									
	m	RMAD		m	RMAD		m	RMAD	
	63	0.173	(0.124)	95	0.146	(0.102)	157	0.145	(0.107)
	70	0.132	(0.090)	105	0.109	(0.080)	172	0.103	(0.082)
	77	0.123	(0.081)	116	0.099	(0.070)	188	0.086	(0.062)
	84	0.133	(0.087)	126	0.112	(0.077)	204	0.087	(0.058)
	91	0.151	(0.089)	136	0.139	(0.081)	219	0.101	(0.062)

Table 1: Results of the Monte Carlo simulation for an ARFIMA(1,H = 0.9,0) model with $a_1 = 0.5$. Standard errors of the relative mean absolute deviation (RMAD) are in parenthesis.

	$T = 400$			$T = 600$			$T = 1000$	
Beta kernel:								
b	RMAD		b	RMAD		b	RMAD	
0.08	0.292	(0.160)	0.08	0.250	(0.132)	0.08	0.218	(0.102)
0.115	0.178	(0.106)	0.115	0.144	(0.078)	0.115	0.119	(0.053)
0.15	0.130	(0.067)	0.15	0.113	(0.048)	0.15	0.105	(0.039)
0.185	0.127	(0.058)	0.185	0.129	(0.057)	0.185	0.142	(0.056)
0.22	0.148	(0.068)	0.22	0.167	(0.070)	0.22	0.194	(0.061)
Semiparametric:								
m	RMAD		m	RMAD		m	RMAD	
91	0.554	(0.183)	126	0.710	(0.192)	62	0.867	(0.255)
99	0.535	(0.174)	136	0.703	(0.184)	78	0.858	(0.242)
106	0.517	(0.165)	146	0.694	(0.176)	94	0.856	(0.235)
113	0.504	(0.160)	156	0.686	(0.168)	110	0.860	(0.229)
120	0.487	(0.150)	166	0.676	(0.163)	125	0.863	(0.223)

Table 2: Results of the Monte Carlo simulation for an ARFIMA(1,$H = 0.6$,0) model with $a_1 = 0.5$. Standard errors of the relative mean absolute deviation (RMAD) are in parenthesis.

	$T = 400$			$T = 600$			$T = 1000$	
Beta kernel:								
b	RMAD		b	RMAD		b	RMAD	
0.0011	0.9429	(0.396)	0.0062	1.3882	(0.428)	0.0415	1.8211	(0.341)
0.0016	0.9383	(0.394)	0.0071	1.3874	(0.422)	0.0432	1.8094	(0.338)
0.0021	0.9381	(0.391)	0.0076	1.3872	(0.419)	0.0449	1.7974	(0.334)
0.0027	0.9395	(0.388)	0.0085	1.3875	(0.413)	0.0466	1.7852	(0.330)
0.0032	0.9419	(0.384)	0.0095	1.3886	(0.408)	0.0483	1.7727	(0.327)
Semiparametric:								
m	RMAD		m	RMAD		m	RMAD	
63	1.401	(0.475)	85	1.792	(0.605)	125	2.440	(0.698)
70	1.354	(0.451)	95	1.693	(0.589)	141	2.243	(0.712)
77	1.310	(0.434)	105	1.656	(0.568)	157	2.125	(0.688)
84	1.344	(0.409)	116	1.719	(0.542)	172	2.140	(0.675)
91	1.345	(0.381)	126	1.810	(0.512)	188	2.230	(0.655)

Table 3: Results of the Monte Carlo simulation for an ARFIMA(2,$H = 0.9$,0) model with $a_1 = 1.172$, $a_2 = -0.707$. Standard errors of the relative mean absolute deviation (RMAD) are in parenthesis.

Table 1 reports the results for an ARFIMA(1,$H = 0.9$,0) model with $a_1 = 0.5$. In that situation, the semiparametric estimator provides the best results whatever the sample size is. The corresponding value of m varies with the sample size; the ratio between m and the sample size is around $\lambda_m \approx 0.19$. Note that for $T = 1000$, the adaptive value of m found in Delgado and Robinson (1996a) converges to 81 (in the conventions of the latter, it corresponds to the frequency $\lambda_m = (2\pi) \times 81/1000 \approx 0.51$). The contrast with the optimal value of m found here is explained by our different objective function: whereas Delgado and Robinson (1996a) concentrates on the mean square error, we consider the RMAD.

In Table 2 we consider the same process except that $H = 0.6$, that is our simulated time series still has a long range dependence, but now with a memory that is "shorter". In that situation, the Beta kernel shows a dramatic improvement when it is compared to the semiparametric estimator. This was expected, because one of our motivations in introducing the Beta kernel is its adaptivity to the memory of the time series.

Another strongly dependent process with $H = 0.9$ is considered in Table 3, however with a more complex dynamical structure. As it is showed in Figure 4(b), the spectral density of that process is not monotone and presents a cycle between frequencies 0.1 and 0.2. The semiparametric estimator is not well-fitted to that situation of non monotone spectrum, as it is confirmed by the results of the Monte Carlo simulation. In contrast, the performance of the Beta kernel is better and demonstrates the good finite sample behavior of the estimator outside the origin. The spectral density of this ARFIMA(2,$H = 0.9$,0) appeared to be very difficult to estimate and it was not straightforward to select the bandwidth of the Beta kernel estimator. In the next section, we give a fully adaptive estimator computed with a data-driven bandwidth b.

4 Empirical results

4.1 Data-driven choice of the bandwidth parameter

The selection of the bandwidth parameter from the data is a relevant question that is addressed in the literature. In our empirical exercise below, we use the generalized leave-one-out spectral technique of Hurvich (1980). In that approach, the function

$$I_T^{-j}(\omega_k) = \begin{cases} I_T(\omega_k) & k \neq j \\ \{I_T(\omega_{j-1}) + I_T(\omega_{j+1})\}/2 & k = j \end{cases}$$

is defined for each $j = 1, \ldots T$. The Beta kernel smoothing with bandwidth b is applied to $I_T^{-j}(\omega_k)$ and is denoted $\hat{f}_b^{-j}(\lambda)$. The cross-validation is motivated by the approximate independence between $\hat{f}_b^{-j}(\omega_j)$ and $I_T(\omega_j)$. In our context of estimation under the L^1 loss, it takes the following form:

$$CV(b) = \sum_{j \in \mathcal{J}} |\hat{f}_b^{-j}(\omega_j) - I_T(\omega_j)| \tag{4.1}$$

		$T=400$		$T=600$		$T=1000$	
ARFIMA(1,$H=0.9$,0) model:	$\hat{b}_{cv}$	0.118	(0.094)	0.107	(0.090)	0.089	(0.086)
	RMAD$_{\mathcal{J}}$	0.503	(0.320)	0.504	(0.354)	0.549	(0.361)
	RMAD$_{\circ}$	1.466	(0.421)	1.412	(0.413)	1.340	(0.384)
ARFIMA(1,$H=0.6$,0) model:	$\hat{b}_{cv}$	0.421	(0.125)	0.411	(0.131)	0.394	(0.139)
	RMAD$_{\mathcal{J}}$	0.335	(0.132)	0.370	(0.128)	0.400	(0.154)
	RMAD$_{\circ}$	1.056	(0.148)	1.045	(0.139)	1.039	(0.109)
ARFIMA(2,$H=0.9$,0) model:	$\hat{b}_{cv}$	0.280	(0.199)	0.250	(0.200)	0.213	(0.196)
	RMAD$_{\mathcal{J}}$	0.537	(0.419)	0.500	(0.419)	0.584	(0.398)
	RMAD$_{\circ}$	13.291	(8.090)	12.203	(8.233)	10.791	(8.16)

Table 4: The performance of the adaptive Beta kernel estimator from 1000 Monte Carlo simulations on the three ARFIMA models. The line $\hat{b}_{cv}$ gives the averages and the standard deviations of the adaptive bandwidth. The line RMAD$_{\mathcal{J}}$ gives the averages and s.d. of the RMAD adaptive estimator over $J = [T^{1/5}, (T/2) - \sqrt{T}]$. The line RMAD$_{\circ}$ gives the same statistic over all discrete frequencies in $(0, 0.5)$.

where $\mathcal{J}$ denotes a given discrete range of frequencies.

In order to evaluate the performance of $CV(b)$ for the choice of the bandwidth, Table 4 reports the results of a Monte Carlo simulation on the three ARFIMA models given in Figure 4. For each sample size T, the bandwidth minimizing (4.1) is found and the table gives the average and standard deviation of the selected bandwidths over 1000 simulations. As expected the adaptive bandwidth is decreasing as the sample size increases. In the simulations, the set $\mathcal{J}$ is choosed to be 100 equidistant points in the interval $\mathcal{J} = [T^{1/5}, \frac{T}{2} - \sqrt{T}]$. For each simulated time series, we focus on the adaptive estimator, that is the Beta kernel estimator computed at the bandwidth minimizing the Cross Validation (4.1). In Table 4 we also estimate the error of the adaptive estimator. The measure of the error considered here is the RMAD computed over $\mathcal{J}$ (denoted by RMAD$_{\mathcal{J}}$ in Table 4), and the RMAD computed over all discrete frequencies in the interval $(0, 0.5)$ (denoted by RMAD$_{\circ}$ in the table).

The deviation found by RMAD$_{\circ}$ is of course larger than the one based on RMAD$_{\mathcal{J}}$ because the bandwidth was optimized on frequencies $\mathcal{J}$. Because RMAD$_{\circ}$ is computed over a fixed range of frequencies $(0, 0.5)$, it is comparable over sample size and Table 4 shows the improvement of the estimator with that respect.

4.2 Nonparametric analysis of S&P 500

We end this study by an application of the data driven estimator on the absolute value of the log returns of the S&P 500 index. In Section 2.2, we already show the use of the estimator on some paths of the stock price that were analysed in Lobato and Savin (1998). Below we consider the path between January 1973 and December 1994, but we also consider two more recent segments of data: from January 1995 and December 2001, and from January 2002 and May 2009.

In Figure 5 we superimpose the logarithm of the data-driven estimator obtained from the three periods of time. In each segment of time, the data are standardised by their standard deviation for the sake of comparison. The bandwidth that is selected by the Cross-validation method is $b = 1.514 \times 10^{-4}$ for the period 1973–1994, $b = 7.475 \times 10^{-4}$ for the period 1995–2001, and $b = 2.815 \times 10^{-5}$ for the period 2002–2009. Nota that the Cross-validation do not provide a clear minimum for the period 1995–2001 because it is flat for $b > 7.475 \times 10^{-4}$.

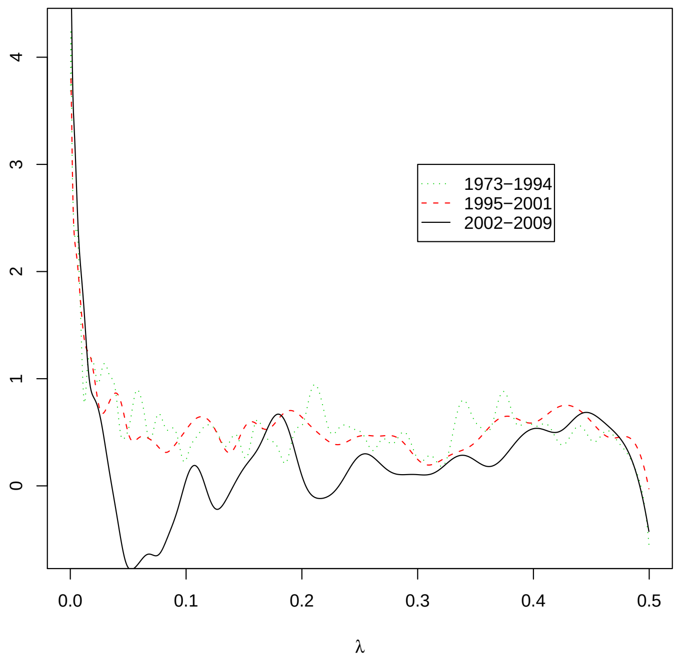


Figure 5: The data-driven log-spectral estimator of the standardized absolute value of the S&P 500 log returns is superimposed for three different periods of time.

From the estimation, it apparent that the spectrum over periods 1973–1994 and 1995–2001 shows clear similarities, whereas the most recent data shows a different behavior for low frequencies. Beyond the frequency zero, the spetrum shows local minima corresponding to various periodocities in the absolute returns. Some periodicities are coherent between the three segments of time. Because the Beta kernel spectral estimator is consistent whatever is the memory of the time series, this empirical example shows that it might be a valuable ingredient in the economic study of the hidden seasonality stock prices.

A Appendix: Proofs

In the proofs, we denote by $K(\cdot, \alpha, \beta)$ the probability density of a Beta(α, β) random variable. The first lemma establishes the uniform convergence of the bias of the Beta kernel estimator.

LEMMA A.1. *If the spectral density f is a continuous function on the interval $(0,1)$, then for any compact I in $(0,1)$, the Beta kernel estimator (2.1) is such that*

$$\sup_{\lambda \in I} \left| I\!E(\hat{f}(\lambda)) - f(\lambda) \right| \longrightarrow 0 \qquad \text{as } T \to \infty$$

provided that $b = b(T) \to 0$.

Proof. We start by recalling a useful property of Beta distributions. If μ_λ and σ_λ^2 denote respectively the mean and the variance of the random variable Z where Z has a Beta$\{1 + \lambda/b, 1 + (1-\lambda)/b\}$ distribution, then there exists a positive, finite constant M such that $\mu_x = \lambda + b(1-2\lambda) + \Delta_1(\lambda)$, $\sigma_\lambda^2 = b\lambda(1-\lambda) + \Delta_2(\lambda)$ and $|\Delta_j(\lambda)| \leqslant Mb^2$ for $j = 1$ and 2 (see e.g. Johnson, Kotz, and Balakrishnan (2000)).

To prove the result, we first consider the approximation $I\!E(\hat{f}(\lambda))$ by $\int_0^1 K_{b,\lambda}(u) f(u) du$ given in (3.1). Consider the following decomposition of the dominant term in (3.1):

$$\begin{aligned}
&\left| \int_0^1 \{f(t) - f(\lambda)\} K\{t, \lambda/b + 1, (1-\lambda)/b + 1\} dt \right| \\
&\quad \leqslant \int_{|t-\mu_\lambda|<\delta} |f(t) - f(\lambda)| K\{t, \lambda/b + 1, (1-\lambda)/b + 1\} dt + \int_{\mu_\lambda+\delta}^1 (\ldots) + \int_0^{\mu_\lambda-\delta} (\ldots) \\
&\quad = \mathrm{I} + \mathrm{II} + \mathrm{III}.
\end{aligned}$$

and we now show the convergence to zero of the three terms.

Since f is uniformly continuous on I, for any $\epsilon > 0$ there exists a $\delta > 0$ such that $|f(t) - f(\lambda)| < \epsilon$ for $|\lambda - t| < \delta$. Therefore $\mathrm{I} \leqslant \epsilon$ for all $b \leqslant b_\epsilon^{\mathrm{I}}$.

Using Chebyshev's inequality and the above bound for σ_λ^2 we also get

$$\begin{aligned} \text{II} &\leqslant \frac{2}{\delta^2} \sup_{t>\mu_\lambda+\delta} |f(t)|\sigma_\lambda^2 \\ &\leqslant \frac{1}{2\delta^2} \sup_{t>\mu_\lambda+\delta} |f(t)| (b + 4Mb^2) \\ &\leqslant \epsilon \end{aligned}$$

for all $b \leq b_\epsilon^{\text{II}}$.

To address the convergence of III we assume without loss of generality that $f(t) > f(\lambda)$ and that $f(t) \sim t^{-\beta}$ for t close to the origin . If ξ denotes the Beta$\{\lambda/b-\beta+1, (1-\lambda)/b+1\}$ random variable, we then write

$$\begin{aligned} \text{III} &\leqslant 2\int_0^{\mu_\lambda-\delta} t^{-\beta} K\{t, \lambda/b+1, (1-\lambda)/b+1\} dt \\ &= \frac{2B\{\lambda/b-\beta+1, (1-\lambda)/b+1\}}{B\{\lambda/b+1, (1-\lambda)/b+1\}} \int_0^{\mu_\lambda-\delta} K\{t, \lambda/b-\beta+1, (1-\lambda)/b+1\} dt \\ &\leqslant \frac{2B\{\lambda/b-\beta+1, (1-\lambda)/b+1\}}{B\{\lambda/b+1, (1-\lambda)/b+1\}} \frac{\mathbb{V}\text{ar}(\xi)}{\delta^2} \\ &\leqslant \epsilon \end{aligned}$$

for $b \leqslant b_\epsilon^{\text{III}}$, and because it is easy to show that $\mathbb{V}\text{ar}(\xi) = b\lambda(1-\lambda) + O(b^2)$.

Combining the three convergence that have been proved we get $\sup_{x\in[0,1]} |\mathbb{E}\{f_b(x)\} - f(x)| < 3\epsilon$ for all $b \leqslant \min(b_\epsilon^{\text{I}}, b_\epsilon^{\text{II}}, b_\epsilon^{\text{III}})$. □

Proof of Proposition 3.1. Since Lemma A.1 establishes a sufficient control of the bias term, it remains to prove the weak convergence of the variation term $\sup_{\lambda\in I} |\hat{f}(\lambda) - \mathbb{E}(\hat{f}(\lambda))|$. Without loss of generality, we suppose that $I = [\eta_1, \eta_2]$ where $0 < \eta_1 < \eta_2 < 1$.

The derivative with respect to $\lambda \in I$ of the beta kernel is given by

$$\frac{dK_{b,\lambda}(t)}{d\lambda} = \frac{1}{b} K_{b,\lambda}(t) \left\{ \ln\left(\frac{t}{1-t}\right) + \psi\left(\frac{1-\lambda}{b}+1\right) - \psi\left(\frac{\lambda}{b}+1\right) \right\}$$

where ψ is the digamma function and satisfies $\psi(x+1) = \ln(x) + (2x)^{-1} - \sum_{j=1}^{\infty} (2j\, x^{2j})^{-1} B_{2j}$ with B_{2j} beeing Bernoulli numbers (see Abramowitz and Stegun (1972) for more details). Also, from Chen (2000) there exists a positive, finite constant c_1 such that $K_{b,\lambda}(t) \leqslant c_1 b^{-1/2}\{\lambda(1-\lambda)\}^{-1/2}$. We conclude that,

$$\begin{aligned} \left|\frac{dK_{b,\lambda}(t)}{d\lambda}\right| &= \frac{1}{b} K_{b,\lambda}(t) \left| \ln\left(\frac{t}{1-t}\right) + \ln\left(\frac{1-\lambda}{\lambda}\right) + \frac{b}{2}\left(\frac{1}{1-\lambda} - \frac{1}{\lambda}\right) + O(b^2) \right| \\ &\leqslant \frac{C}{b^{3/2}} \end{aligned}$$

for some constant C depending on η_1 and η_2. Therefore, for λ and $\lambda \in I$ we can write

$$\begin{aligned} |\hat{f}(\lambda) - \hat{f}(\lambda')| &= \frac{1}{T}\sum_{j=1}^{T} |K_{b,\lambda}(\omega_j) - K_{b,\lambda'}(\omega_j)| I_T(\omega_j) \\ &\leqslant \frac{C}{b^{3/2}T}|\lambda - \lambda'| \sum_{j=1}^{T} I_T(\omega_j). \end{aligned}$$

Hence, if we control as above the smaller order approximation terms in the expectation,

$$\begin{aligned} |\mathbb{E}\hat{f}(\lambda) - \mathbb{E}\hat{f}(\lambda')| &\leqslant \mathbb{E}|\hat{f}(\lambda) - \hat{f}(\lambda')| \\ &\leqslant \frac{C}{b^{3/2}}|\lambda - \lambda'|\{\gamma(0) + o(1)\} \end{aligned}$$

Let $\epsilon > 0$ and consider a partition of the interval $[\eta_1, \eta_2]$ into $N = [b^{-\epsilon-3/2}]$ subintervals $\{I_j\}$ of equal length, with center λ_j. Then

$$\sup_{\lambda \in I_j} |\hat{f}(\lambda) - \mathbb{E}\hat{f}(\lambda)| \leqslant |\hat{f}(\lambda_j) - \mathbb{E}\hat{f}(\lambda_j)| + \frac{C}{Nb^{3/2}}\{\gamma(0) + o(1)\}$$

Therefore,

$$\sup_{\lambda \in I} |\hat{f}(\lambda) - \mathbb{E}\hat{f}(\lambda)| \leqslant \max_{1 \leqslant j \leqslant N} |\hat{f}(\lambda_j) - \mathbb{E}\hat{f}(\lambda_j)| + \frac{C}{Nb^{3/2}}\{\gamma(0) + o(1)\}.$$

Using (3.2) and the Chebychev inequality, we also note that $\hat{f}(\lambda) - \mathbb{E}\hat{f}(\lambda) = O_P(b^{-1/2}T^{-1})$ for all $\lambda \in I$, and therefore $\max_j |\hat{f}(\lambda_j) - \mathbb{E}\hat{f}(\lambda_j)| = O_P(Nb^{-1/2}T^{-1}) = O_P\left(b^{-2-\epsilon}T^{-1/2}\right)$ which gives the result. □

Proof of Proposition 3.2. The divergence of the spectral density at the origin implies that for any $C > 0$ there exists $\delta > 0$ such that $f(t) > C$ for all $t < \delta$. We first show that the expectation of the Beta kernel estimator diverges at frequency zero when there is a pole at the origin of the spectrum. In (3.1) we have computed the expectation for $\lambda \neq 0$; the situation is slightly different at $\lambda = 0$. Still, we can write that

$$\mathbb{E}(\hat{f}(0)) = \int_0^1 K_{b,0}(u) f(u) du + R_1 + R_2$$

where $R_1 = O(T^{-1})$. To evaluate R_2 we note that $K_{b,0}(\omega_j) = b^{-1}(1+b)(1-j/T)^{1/b}$ which is bounded by $b^{-1}(1+b)$ and then we can apply the arguments of Robinson (1994b) on the pooled periodogram in order to show that $R_2 = O(b^{-1}(m_T/T)^{1-\beta}) = o(1)$ under the constraint (ii) of the proposition. Therefore, using that the spectrum f is integrable,

$$\begin{aligned}\mathbb{E}(\hat{f}(0)) &= b^{-1}(1+b)\int_0^1 (1-t)^{1/b} f(t)dt + o(1)\\ &> b^{-1}(1+b)C\int_0^{\delta}(1-t)^{1/b}dt + o(1)\\ &> C(1-(1-\delta)^{1/b+1}) + o(1).\end{aligned}$$

The first term of the last expression converges to C as b tends to zero, which proves the divergence of $\mathbb{E}(\hat{f}(0))$. To show the convergence in probability, we again use the Chebychev inequality and (3.2): for any $\epsilon > 0$ and for a sequence λ such that $\lambda/b \to \kappa$,

$$P(|\hat{f}(\lambda) - \mathbb{E}\hat{f}(\lambda)| > \epsilon) = O\left(\frac{1}{Tb^{1+2\beta}}\right)$$

which proves the announced result. □

Proof of Corollary 3.1. Using again that $K_{b,0}(\omega_j) = b^{-1}(1+b)(1-j/T)^{1/b}$ we can write

$$|\mathbb{E}(\hat{f}(0)) - f(0)| \leqslant \frac{b+1}{b}\int_0^1 (1-t)^{1/b}|f(t)-f(0)| + o(1).$$

Since f is continuous on the right side of 0, for any $\epsilon > 0$ there exists a $\delta > 0$ such that $|f(t) - f(0)| < \epsilon$ for $t < \delta$. Splitting the integral over $[0,1]$ in $[0,\delta]\cup[\delta,1]$, we get the bound

$$\epsilon\frac{b+1}{b}[1-(1-\delta)^{1+1/b}] + 2M(1-\delta)^{1+1/b}$$

where $M := \sup_{t\in[0,1]}|f(t)|$. Since $b \to 0$ and the bound holds for every $\epsilon > 0$, we get $|\mathbb{E}(\hat{f}(0)) - f(0)| = o(1)$. Finally, as in the proof Proposition 3.2, we conlude with the Chebychev inequality and (3.2) that lead to $P(|\hat{f}(0) - \mathbb{E}\hat{f}(0)| > \epsilon) = O(b^{-1}T^{-1})$, and get the stated result. □

Proof of Proposition 3.3. We start by proving the relative convergence of the bias term, that is $|\{\mathbb{E}(\hat{f}(\lambda)) - f(\lambda)\}/f(\lambda)| \longrightarrow 0$ as $\lambda/b \to \kappa$. We proceed as in the beginning of the proof of Lemma A.1. Omitting the negligible terms, we use the decomposition

$$\begin{aligned}&\left|\frac{\mathbb{E}(\hat{f}(\lambda)) - f(\lambda)}{f(\lambda)}\right|\\ &\leqslant \int_{|t-\mu_\lambda|<\delta}\frac{|f(t)-f(\lambda)|}{f(\lambda)}K\{t,\lambda/b+1,(1-\lambda)/b+1\}dt + \int_{\mu_\lambda+\delta}^1(\ldots) + \int_0^{\mu_\lambda-\delta}(\ldots)\\ &= \mathrm{I} + \mathrm{II} + \mathrm{III}.\end{aligned}$$

in which I $\leqslant \epsilon$ for all $b \leqslant b_\epsilon^{\mathrm{I}}$ and II $\leqslant \epsilon$ for all $b \leqslant b_\epsilon^{\mathrm{II}}$. The treatment of the term III is not as in Lemma A.1. Assuming that $f(t) > f(\lambda)$ for $t \leqslant \mu_\lambda + \delta$ without loss of generality, we can write by Assumption 3.1 that

$$\frac{|f(t) - f(\lambda)|}{f(\lambda)} \sim \frac{|t^{-\beta} - \lambda^{-\beta}|}{\lambda^{-\beta}} \leqslant 1$$

and therefore

$$\begin{aligned} \mathrm{III} &\leqslant \int_0^{\mu_\lambda - \delta} K\{t, \lambda/b + 1, (1-\lambda)/b + 1\} dt \\ &\leqslant \frac{\mathbb{V}\mathrm{ar}(\xi)}{\delta^2} \leqslant \epsilon \end{aligned}$$

for all $b \leqslant b_\epsilon^{\mathrm{III}}$, where ξ is a Beta$\{\lambda/b - \beta + 1, (1-\lambda)/b + 1\}$ random variable which is such that $\mathbb{V}\mathrm{ar}(\xi) = b\lambda(1-\lambda) + O(b^2)$. By combining the three terms, the bias term is bounded by 3ϵ for all $b \leqslant \min(b_\epsilon^{\mathrm{I}}, b_\epsilon^{\mathrm{II}}, b_\epsilon^{\mathrm{III}})$.

Finally, we control the convergence of the variation term using the Chebychev's inequality. Indeed for λ such that $\lambda/b \to \kappa$

$$\begin{aligned} P\left(\frac{|\hat{f}(\lambda) - \mathbb{E}(\hat{f}(\lambda))|}{f(\lambda)} > \epsilon\right) &\leqslant \frac{\mathbb{V}\mathrm{ar}(\hat{f}(\lambda))}{f(\lambda)^2 \epsilon^2} \\ &= \frac{C(\kappa)}{Tbf(\lambda)^2\epsilon^2}\{f(\lambda)^2 + O(T^{-1})\} \\ &= O_P(b^{-1}T^{-1}), \end{aligned}$$

which implies the weak convergence of the variation term, and therefore ends the proof. □

References

Abramowitz, M., and I.-A. Stegun (1972): *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables.* Dover.

Andrews, D. W., and Y. Sun (2004): "Adaptive Local Polynomial Whittle Estimation of Long-range Dependence," *Econometrica*, 72, 569–614.

Brillinger, D. R. (2001): *Time series: data analysis and theory.* SIAM.

Brown, B., and S. X. Chen (1999): "Beta-Bernstein smoothing for regression curves with compact supports," *Scand. J. Statist.*, 26, 47–59.

Chen, S. X. (1999): "Beta kernel estimators for density functions," *Comput. Statist. Data. Anal.*, 31, 131–145.

——— (2000): "Beta kernel smoothers for regression curves," *Statist. Sinica*, 10, 73–91.

Delgado, M., and P. M. Robinson (1996a): "Optimal spectral bandwidth for long memory," *Statist. Sinica*, 6, 97–112.

——— (1996b): "Optimal spectral kernel for long-range dependent time series," *Statistics and Probability Letters*, 30, 37–43.

Doukhan, P., G. Oppenheim, and M. S. Taqqu (eds.) (2003): *Theory of Long Range Dependence*. Birkhäuser, Birkhuser.

Henry, M. (2007): "Bandwidth Choice, Optimal Rates and Adaptivity in Semiparametric Estimation of Long Memory," in *Long Memory in Economics*, ed. by G. Teyssière, and A. P. Kirman.

Hurvich, C. M. (1980): "Data-Driven Choice of a Spectrum Estimate: Extending the Applicability of Cross-Validation Methods," *J. Amer. Statist. Assoc.*, 14, 227–238.

Johnson, N. L., S. Kotz, and N. Balakrishnan (2000): *Continuous Univariate Distributions*. Wiley-Interscience, 2nd edn.

Lobato, I., and P. M. Robinson (1998): "A nonparametric test for $I(0)$," *Review of Economic Studies*, 65, 475–495.

Lobato, I., and N. Savin (1998): "Real and spurious long memory properties of stock market data," *J. Bus. Econ. Statist.*, 16, 261–283.

Lobato, I., and C. Velasco (2000): "Long Memory in Stock-Market Trading Volume," *J. Bus. Econ. Statist.*, 18, 410–427.

Moulines, E., and P. Soulier (1999): "Broadband log-periodogram regression of time series with long-range dependence," *Ann. Statist.*, 27, 1415–1439.

Ohanissian, A., J. R. Russel, and R. A. Tsay (2008): "True of Spurious Long Memory? A New Test," *J. Bus. Econ. Statist.*, 26, 161–175.

Palma, W. (2007): *Long-Memory Time Series: Theory and Methods*, Wiley Series in Probability and Statistics. Wiley.

Robinson, P. M. (1994a): "Rates of convergence and optimal spectral bandwidth for long range dependence," *Probab. Theory Related Fields*, 99, 443–473.

——— (1994b): "Semiparametric analysis of long-memory time series," *Ann. Statist.*, 22, 515–539.

——— (1995): "Gaussian semiparametric estimation of long range dependence," *Ann. Statist.*, 23, 1630–1661.

——— (ed.) (2003): *Time Series with Long Memory*. Oxford University Press.

Robinson, P. M., and M. Henry (2003): "Higher-order Kernel Semiparametric M-estimation of Long Memory," *J. Econometrics*, 114, 1–27.