

# Nonparametric Beta Kernel Estimator for Long and Short Memory Time Series

May 20, 2019

## Abstract

This paper introduces a new nonparametric estimator of the spectral density that is given by smoothing the periodogram using Beta kernel density. 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 show that the proposed estimator automatically adapts to the long- and the short-range dependency of the process. A cross-validation procedure is 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 performance of the Beta kernel estimator.

keywords: Beta kernel smoothing; Cross-validation; Long range dependence; Nonparametric estimation; Periodogram; Short memory; Spectral density.

## 1 Introduction

The estimation of a spectral density often requires the knowledge of whether a stationary time series is short or long memory. Short memory time series is well known by its autocovariance function that decreases rapidly as the time difference increases. In the case of long memory

time series, there is much stronger dependence between values at different times and the decay of the autocovariance function is slow. Long memory or long range dependence is characterized by a spectral density that is unbounded at frequency zero, see Robinson (2003) and Palma (2007). Therefore, the choice of an optimal nonparametric estimator will be different if the spectral density is bounded or not. It is our goal to go beyond that limitation and propose an estimator that is applicable to any stationary data, being long or short range dependent.

A well-established nonparametric estimation procedure consists of estimating the parameter  $d_0$  of the long memory process, where the spectral density  $f$  is assumed to behave like

$$f(\lambda) = |\lambda|^{-2d_0}L(\lambda) \tag{1}$$

as  $\lambda \rightarrow 0+$ , for  $d_0 \in (0, 1/2)$  and  $L(\lambda)$  is a slowly varying function with  $0 < L(0) < \infty$ . The estimation of  $d_0$  was studied by Robinson and Henry (2003), Andrews and Sun (2004) and Henry (2007), among others. See also the recent surveys by Doukhan et al. (2003), Robinson (2003) and Palma (2007). Inference on  $d_0$  allows us to test whether  $d_0$  is significantly larger than zero if the process is long memory (see Lobato and Robinson (1998), Lobato and Velasco (2000), Ohanissian et al. (2008)). The testing step is important because the asymptotic distribution of the spectral density estimator is usually not the same for  $d_0 = 0$  and  $d_0 > 0$ . If the process is short memory, the nonparametric estimation of its spectral density becomes a classical problem of inference. Ferreira and Rodriguez-Poo (1998), for instance, studied this problem by proposing symmetric kernel smoothing of the periodogram. If the time series is long memory, Robinson (1995) among others proposed to estimate the spectrum for  $\lambda$  close to zero by  $\hat{C}|\lambda|^{-2\hat{d}_0}$  for a consistent estimator of  $d_0$  and  $C$ . Away from the origin, a classical symmetric kernel can be used in order to evaluate the spectrum for  $\lambda$  away for zero.

Different studies employed standard kernel smoothing to estimate the spectral density. For instance, see Riedel and Sidorenko (1996), Lee (1997, 2001), Fan and Kreutzberger (1998), Ferreira and Rodriguez-Poo (1998), Hannig and Lee (2004) and Birr et al. (2017), among others. Such approach, however, is known to not account for the boundaries bias. Hence, using the standard kernel methods for the spectral density are not recommended

because they will provide a severe bias at and near zero. In this paper, we propose a new estimator for the spectral density based on Beta kernel smoothing of the periodogram. Our methodology is nonparametric and can be applied for processes not necessarily Gaussian. The Beta kernel is a non-symmetric density function with a shape varying according to the frequency where the spectrum is estimated. Beta kernel smoothing was initially introduced by Brown and Chen (1999) for estimating a compactly supported regression curve. It has then been used to address the boundary bias problem for regression and density functions estimation (Chen (1999, 2000)). Because 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 dependence range 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 introduce the Beta kernel estimator of the spectral density. The asymptotic properties of the estimator and the bandwidth selection are discussed in Section 3. First, we study the behaviour of the estimator away from the origin and establish its uniform convergence on any compact set of frequencies. Second, we show that the estimator is unbounded in probability at the origin for long range time series, and is bounded for short memory processes. Third, we derive a stronger result by establishing the relative convergence of the estimator at the origin. Finally, we study a cross-validation method for selecting the smoothing parameter following the general method of Hurvich (1980). In Section 4, we study the finite sample properties of the Beta kernel estimator. By considering three parametric (FARIMA) models, we show the adaptation of the proposed estimator to the range of memory of the process and we compare the empirical performance of our estimator with the semiparametric estimator of Robinson (1995). The Beta kernel estimator of the spectral density is then applied for analyzing the returns and absolute returns of the S&P500 index and shows that the proposed estimator detects the long- (resp. short-) range dependence of the absolute returns (resp. the return) time series. Appendix A contains the proofs of the theoretical results.

## 2 The Beta kernel estimator of the spectral density

Our goal is to develop a kernel smoothing method for estimating the spectral density for a short or long memory time series that is free of boundary bias. This 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  $\lambda = 0$ . Suppose we observe  $X_1, \dots, X_T$  from a stationary process spectral density

$$f(\lambda) = \sum_{k \in \mathbb{Z}} \gamma(k) \exp(-2\pi i \lambda k), \lambda \in (0, 1),$$

where  $\gamma(k) = Cov(X_t, X_{t+k})$  is the covariance function that satisfies one of the statements,  $(\star) \gamma \in L_1$  or  $(\star\star) \lim_{k \rightarrow \infty} \gamma(k)k^\alpha = c$  for some constants  $0 < \alpha < 1$  and  $c > 0$ . For more details about the two statements, see Gubner (2005) and Proposition 2.3 in Hu et al. (2015). If the covariance function satisfies  $(\star)$ , then the spectral density  $f(\lambda)$  exists and is a nonnegative bounded function. Therefore, we have a time series with short-range dependence. Whereas, if  $(\star\star)$  is fulfilled, then the spectral density exists and satisfies  $\lim_{\lambda \rightarrow 0} |\lambda|^{1-\alpha} f(\lambda) = c_f$  for some constant  $c_f > 0$ , and therefore we have a long memory time series.

For the sake of simplicity, we assume the stationary process to be zero mean. The periodogram defined by :

$$I_T(\omega_j) = \frac{1}{T} \left| \sum_{t=1}^T X_t \exp(-2\pi i \omega_j t) \right|^2, \quad \omega_j = \frac{j}{T} \text{ for } j = 1, 2, \dots, T, \quad (2)$$

is known to be an asymptotically unbiased and inconsistent estimator of the spectral density  $f$ , for details see Brillinger (2001). However, one can derive a consistent estimator by using appropriate smoothing of  $I_T$  over frequencies  $\omega_j$ . For this purpose, we propose the estimator

$$\hat{f}(\lambda) = \frac{1}{T} \sum_{j=1}^T K_{b,\lambda}(\omega_j) I_T(\omega_j), \quad (3)$$

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

$$K_{b,\lambda}(\omega) = \frac{\omega^{\lambda/b} (1-\omega)^{(1-\lambda)/b}}{B\left(\frac{\lambda}{b} + 1, \frac{1-\lambda}{b} + 1\right)} \mathbb{I}_{\{0 \leq \omega \leq 1\}}, \quad (4)$$

with  $B$  the Beta function,  $b$  a real nonnegative real value playing the role the smoothing parameter, and  $\mathbb{I}(\cdot)$  the indicator function. The Beta kernel,  $K_{b,\lambda}$ , is the density function of a Beta distribution with parameters  $1 + \lambda/b$  and  $1 + (1 - \lambda)/b$ .

In contrast to most kernel estimators, the estimator  $\hat{f}(\lambda)$  does not use a symmetric kernel but a kernel whose shape varies with  $\lambda$ . As noticed by Chen (1999), the variance of the Beta random variable with parameters  $1 + \lambda/b$  and  $1 + (1 - \lambda)/b$  is of order

$$b\lambda(1 - \lambda) + O(b^2). \tag{5}$$

Hence, the amount of smoothing given in (5) is small at the border of the support and tends to zero when  $b$  converges to zero.

### 3 Properties of the estimator

In this section, we explore the asymptotic properties of the estimator and a data-driven bandwidth parameter of the proposed estimator. In the sequel, we use the notation  $h_1(\lambda) \sim h_2(\lambda)$ , for some functions  $h_1$  and  $h_2$ , if there exists two nonnegative finite constant  $c_1$  and  $c_2$  such that  $c_1 \leq h_1(\lambda)/h_2(\lambda) \leq c_2$  for all  $\lambda$ . We suppose that the following assumption is satisfied:

**Assumption 1** *The spectral density  $f$  is such that  $f(\lambda) \sim \lambda^{-\beta}$  as  $\lambda \rightarrow 0^+$ , for  $0 \leq \beta < 1$ . For  $\lambda > 0$ ,  $f$  is a positive and bounded Lipschitz continuous function on  $[\lambda, 1)$ .*

#### 3.1 Behavior of the estimator outside the origin

Given a stationary zero mean time series  $\{X_t; t = 1, \dots, T\}$  with a spectral density  $f(\lambda)$ , we derive below the convergence rate of the bandwidth  $b = b(T)$  for which 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  $(0, 1)$ .

**Proposition 1** *Suppose that  $f$  is twice differentiable and Assumption 1 holds. Then, the expectation of  $\hat{f}$  at frequency  $\lambda \neq 0$  is given by*

$$\mathbb{E}(\hat{f}(\lambda)) = f(\lambda) - b\{(1 - 2\lambda)f'(\lambda) + \frac{1}{2}\lambda(1 - \lambda)f''(\lambda)\} + o\left(b + \frac{1}{T^{1/2}b^{1/4}}\right),$$

where  $f'$  and  $f''$  are the first and second derivatives of  $f$ , respectively.

The variance of  $\hat{f}$  is

$$\mathbb{V}\text{ar}(\hat{f}(\lambda)) = \frac{1}{T\sqrt{b}} \frac{f^2(\lambda)}{2\sqrt{\pi\lambda(1-\lambda)}} + o((Tb^{1/2})^{-1}) + \rho_T$$

if  $\lambda/b \rightarrow \infty$  and  $(1 - \lambda)/b \rightarrow \infty$ , with  $\rho_T = O(\log^r(T)/(bT^{2-2\beta}))$  for some  $r > 0$ , and is equal to

$$\mathbb{V}\text{ar}(\hat{f}(\lambda)) = \frac{C(\kappa)}{Tb} \{f(\lambda)^2 + O(T^{-1})\} + \rho_T$$

if  $\lambda/b \rightarrow \kappa$  or  $(1 - \lambda)/b \rightarrow \kappa$ , with  $\kappa > 0$  a constant and  $C(\kappa) = \Gamma(2\kappa + 1)/ (2^{1+2\kappa}\Gamma(\kappa + 1)^2)$ .

For  $\lambda \neq 0$ , Proposition 1 implies that the Beta kernel estimator is asymptotically unbiased with vanishing variance if  $\beta < 1/2$  and the bandwidth  $b = b(T)$  satisfies

$$b + \frac{1}{T\sqrt{b}} \rightarrow 0 \tag{6}$$

as  $T \rightarrow \infty$  and  $b \rightarrow 0$  (which imply that the bias and variance tend to zero). Also, the optimal convergence rate of the mean integrated squared error of the Beta kernel estimator is  $O(T^{-4/5})$ , which is the well known rate for smoothing nonparametric approaches. The constraint on  $\beta$  imposes that the spectrum is square integrable around the pole, and therefore, the mean square error is invariant to the explicit variation of  $f$  around frequency 0. Following Robinson (1994b), 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  $I \subset (0, 1)$ . This result is valid for short and long memory time series.

**Proposition 2** *Suppose that Assumption 1 is fulfilled with  $\beta \in [0, 1/2)$ . If  $b$  satisfies (6) and  $(b^{2+\epsilon}T) \rightarrow \infty$  as  $T \rightarrow \infty$  for some  $\epsilon > 0$ , then, for any compact set  $I \subset (0, 1)$*

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

The proof of Proposition 2 is given in the appendix. The uniform convergence of the Beta kernel estimator on a compact set  $I$  is valid for short ( $\beta = 0$ ) and long memory ( $\beta \in (0, 1/2)$ ) time series. This implies that our proposed estimator is adaptive for both type of time series.

### 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 when the spectrum is bounded or not at  $\lambda = 0$ .

**Proposition 3** *Suppose that Assumption 1 holds for  $\beta \in [0, 1/2)$ . If  $b$  satisfies (6) and  $T^{1-2\beta}b \rightarrow \infty$  as  $T \rightarrow \infty$ , then*

(i) *if  $\beta \in (0, 1/2)$  (long memory process), we have*

$$\hat{f}(0) \xrightarrow{P} +\infty \quad \text{as } T \rightarrow \infty.$$

(ii) *if  $\beta = 0$  (short memory process), we have*

$$\hat{f}(0) \xrightarrow{P} f(0) \quad \text{as } T \rightarrow \infty.$$

We conclude that the Beta kernel estimator is automatically adapted to the “nature of memory” of the spectral density (long or short range). This result has been already illustrated in Figure 4. Note that the symmetric kernel estimator proposed by Ferreira and Rodriguez-Poo (1998) is not consistent at the pole, even for short memory, when the spectral density is not nul at zero.

Proposition 3 does not provide any information about the closeness of the estimator to the spectral density near the origin. In order to have an idea about this closeness. The next result gives more details about the *relative* convergence of the Beta kernel estimator near the origin.

**Proposition 4** *Suppose Assumption 1 is fulfilled for  $\beta \in [0, 1/2)$ . If the bandwidth parameter satisfies (6) and  $Tb \rightarrow \infty$  as  $T \rightarrow \infty$ , then*

$$\left| \frac{\hat{f}(\lambda)}{f(\lambda)} - 1 \right| \xrightarrow{P} 0, \quad \text{as } T \rightarrow \infty$$

*if  $\lambda \rightarrow 0$  such that  $\lambda/b \rightarrow \kappa$  for some constant  $\kappa > 0$ .*

### 3.3 Data-driven choice of the bandwidth parameter

The selection of the bandwidth parameter of the smoothing nonparametric method is well addressed in the literature for symmetric kernels. Here, we consider the generalized leave-one-out spectral technique of Hurvich (1980) for selecting a practical bandwidth parameter  $b$  of the Beta kernel estimator. Let's define

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

for  $j = 1, \dots, T$ . 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 (CV) is motivated by the approximate independence between  $\hat{f}_b^{-j}(\omega_j)$  and  $I_T(\omega_j)$ . In our context of the  $L_1$  loss, we consider

$$CV(b) = \sum_{j \in \mathcal{J}} |\hat{f}_b^{-j}(\omega_j) - I_T(\omega_j)| \quad (7)$$

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 1 reports the results of a Monte Carlo simulations on the three FARIMA models given in Figure 1. For each sample size  $T$ , the bandwidth minimizing (7) is found and the table gives the average and the standard deviation of the selected bandwidths over 1000 simulations. As expected, the adaptive bandwidth decreases as the sample size increases. In the simulations, the set  $\mathcal{J}$  is chosen to be 100 equidistant points of the interval  $\mathcal{J} = [T^{1/5}, \frac{T}{2} - \sqrt{T}]$ . For each simulated time series we focus on the adaptive estimator computed at the bandwidth minimizing (7). In Table 1, we estimated the error of the adaptive estimator. The error measure considered here is the RMAD computed over  $\mathcal{J}$  (denoted by  $\text{RMAD}_{\mathcal{J}}$ ) and the RMAD computed over all discrete frequencies in the interval  $(0, 0.5)$  (denoted by  $\text{RMAD}_{\circ}$ ). The deviation found by  $\text{RMAD}_{\circ}$  is of course larger than the one based on  $\text{RMAD}_{\mathcal{J}}$  because the bandwidth was optimized on frequencies  $\mathcal{J}$ . Because  $\text{RMAD}_{\circ}$  is computed over a fixed range of frequencies  $(0, 0.5)$ , it is comparable over sample size and Table 1 shows the improvement of the estimator with that respect.



	$T = 400$		$T = 600$		$T = 1000$	
FARIMA(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)
FARIMA(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)
FARIMA(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 1: The performance of the adaptive Beta kernel estimator based on 1000 Monte Carlo simulations of three FARIMA models. The row  $\hat{b}_{cv}$  gives the averages and the standard deviations s.d. (inside parentheses) of the adaptive bandwidth. The row RMAD $_{\mathcal{J}}$  gives the averages and s.d. (inside parentheses) of the RMAD adaptive estimator over  $J = [T^{1/5}, (T/2) - \sqrt{T}]$ . The row RMAD $_{\circ}$  gives the same statistic over all discrete frequencies in  $(0, 0.5)$ .

## 4 Empirical results

### 4.1 Finite sample properties

In this section, we examine the properties of the estimator through Monte Carlo simulations. In order to evaluate the estimator quality, we provide a comparison with the semiparametric estimator of Robinson (1994a). The latter semiparametric approach assumes that the spectrum satisfies  $f(\lambda) \sim C_f \lambda^{1-2H}$  as  $\lambda \rightarrow 0^+$ , where  $1/2 < H < 1$ , and proposes consistent estimators for  $H$  and  $C_f$ . Observing that the spectral distribution  $F(\lambda) = \int_0^\lambda f(t)dt$  satisfies

$F(u\lambda)/F(\lambda) \sim u^{2(1-H)}$  for all  $u \in (0, 1)$ , Robinson (1994a) proposed to estimate  $H$  by

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

for a given  $u$  and frequency  $\lambda_m = m/T$ , and since  $F(\lambda) = \frac{C_f}{2-2H} \lambda^{2-2H}$ , he estimated  $C_f$  by

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

Consequently, the spectrum estimator near the origin is  $\hat{f}(\lambda) = \hat{C}_f \lambda^{1-2\hat{H}}$ .

The semiparametric estimator depends on the choice of two parameters,  $u$  and  $m$ . In our computations below, we set  $u = 0.5$  as commonly used in the literature. The choice of  $m$  is however more delicate. Based on the expansion of the asymptotic mean squared error, some rules for the choice of  $m$  have been proposed by Robinson (1994b), and later been refined by Delgado and Robinson (1996a,b).

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

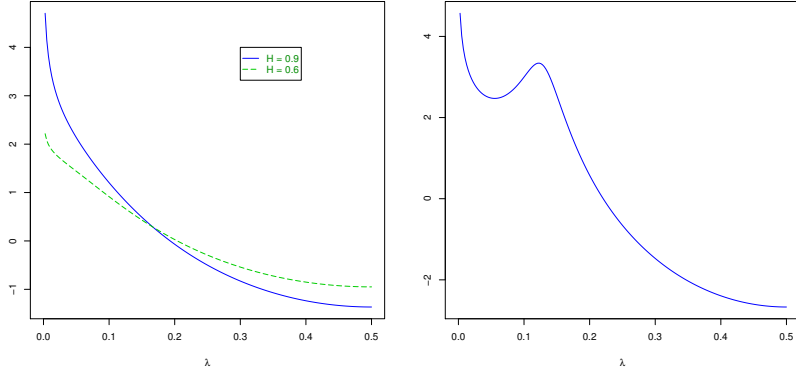
$$f(\lambda) = |1 - \exp(i\lambda)|^{1-2H} h(\lambda), \quad (8)$$

where  $\lambda, H \in [0, 1]$ , and

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

with  $\sigma^2 > 0$ ,  $a(z) = 1 - \sum_{j=1}^p a_j z^j$  and  $b(z) = 1 - \sum_{j=1}^q b_j z^j$ . In Model (8),  $H = 1/2$  corresponds to short memory process if  $0 < h(\lambda) < \infty$ , whereas  $H > 1/2$  corresponds to long memory process.

Figure 1 displays the logarithm of the spectral density of three FARIMA generating models used in the simulation below. The estimators were computed based 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 did not compare it with the Beta kernel estimator over all frequencies but only in a neighbourhood of the frequency zero. Following



(a) FARIMA(1, $H$ ,0) with  $a_1 = 0.5$  and two values of the memory parameter  $H$ .  
 (b) FARIMA(2, $H = 0.9$ ,0) with  $a_1 = 1.172$ ,  $a_2 = -0.707$ .

Figure 1: Logarithm of the three spectral densities of the FARIMA used in the simulations.

Robinson (1994a), we compute the error of estimation on the frequencies  $\omega_j \in (\omega_{j_0}, \omega_{j_1})$ , where  $j_0 = \lceil \sqrt[5]{T} \rceil$  and  $j_1 = \lfloor \sqrt{T} \rfloor$  with  $\lfloor a \rfloor$  the integer part of  $a$ . The computed empirical error is the relative mean absolute deviation

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

where  $\hat{f}_s$  is either the Beta kernel estimator or the seminparametric estimator of Robinson (1994a). Taking the relative mean absolute deviation instead of the mean absolute deviation is driven by the unboundness of  $f$  at frequency zero. Because  $j_0$  and  $j_1$  depend on  $T$ , the range of frequencies where the error is computed is different for different sample sizes. Therefore, the values of RMAD presented in the empirical study below are only comparable for a given sample size.

The FARIMA time series were generated via the library ‘fracdiff’ in R. To avoid the dependence of our conclusions on the choice of the bandwidths, we have computed the RMAD for a range of bandwidths which is  $[0.01, 0.5]$ . The range of  $m$  in the semiparametric estimation is  $[T^{1/2}, T^{4/5}]$ . Tables 2 to 4 display the five results that were the closest to the best RMAD found.

Table 2 reports the results for FARIMA(1, $H = 0.9$ ,0) model with  $a_1 = 0.5$ . The semiparametric estimator provides the best results whatever the sample size is. The corresponding

$T = 400$			$T = 600$			$T = 1000$		
<i>Beta kernel:</i>								
$b$	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)
<i>Semiparametric:</i>								
$m$	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 2: Results of the Monte Carlo simulation for FARIMA(1, $H = 0.9,0$ ) model with  $a_1 = 0.5$ . Standard errors of the relative mean absolute deviation (RMAD) are in parenthesis.

value of  $m$  varies with the sample size; the ratio between  $m$  and the sample size is around  $\lambda_m \approx 0.19$ . For  $T = 1000$ , the adaptive value of  $m$  found in Delgado and Robinson (1996a) converges to 81 which 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 use of the RMAD criterion, whereas Delgado and Robinson (1996a) employed the mean squared error.

In Table 3, we consider the same process except that  $H = 0.6$ , that is, the simulated time series still has a long range dependence but with a shorter memory than the case of  $H = 0.9$ . In that situation, the Beta kernel shows a dramatic improvement compared to the semiparametric estimator, agreeing with the expectation of the adaptivity of the Beta kernel to the memory of the time series.

Another strongly dependent process with  $H = 0.9$  is considered in Table 4 with a more complex dynamical structure. As shown in Figure 1(b), the spectral density of the process is not monotone and presents a cycle between frequencies 0.1 and 0.2. The semiparametric

$T = 400$			$T = 600$			$T = 1000$		
<i>Beta kernel:</i>								
$b$	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)
<i>Semiparametric:</i>								
$m$	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 3: Results of the Monte Carlo simulation for FARIMA(1, $H = 0.6,0$ ) model with  $a_1 = 0.5$ . Standard errors of the relative mean absolute deviation (RMAD) are in parenthesis.

estimator is not well-fitted to that situation of non monotone spectrum, as it is indicated by the results of the Monte Carlo simulation. In contrast, the performance of the Beta kernel is better and demonstrates a good finite sample behavior of the estimator. The spectral density of FARIMA(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.

## 4.2 Nonparametric analysis of S&P 500

We finish this study by an application of the Beta kernel estimator on the absolute value of the log returns of the S&P 500 index. We consider the path between January 1973 and December 1994. Also, we analyse two more recent segments of data: from January 1995 to December 2001 and from January 2002 to May 2009.

In Figure 2, we draw the logarithm of the data-driven estimator obtained from the three periods of time. In each time segment, the data are standardized by their standard deviation

$T = 400$			$T = 600$			$T = 1000$		
<i>Beta kernel:</i>								
$b$	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)
<i>Semiparametric:</i>								
$m$	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 4: Results of the Monte Carlo simulation for FARIMA(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.

for the sake of comparison. The bandwidths selected by the cross-validation method are  $b = 1.514 \times 10^{-4}$  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. The cross-validation does not provide a clear minimum for the period 1995–2001 because it is flat for  $b > 7.475 \times 10^{-4}$ .

Figure 2 shows clear similarities for the spectrum over periods 1973–1994 and 1995–2001, whereas the most recent data shows a different behavior for low frequencies. Beyond the frequency zero, the spectrum shows local minima corresponding to various periodicities in the absolute returns. Some periodicities are coherent between the three segments of time. Because the Beta kernel 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.

An eminent feature of the Beta kernel estimator is its adaptivity to the boundness or

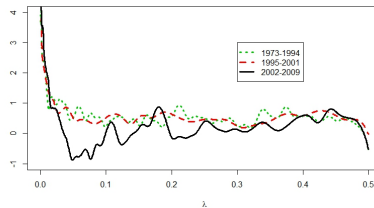


Figure 2: The data-driven log-spectral estimator of the standardized absolute value of the S&P 500 log returns is superimposed for three different periods of time.

unboundness of the spectrum at the origin  $\lambda = 0$ . To illustrate that property, we consider in Figure 3 a segment of the daily absolute returns of the S&P500 that was analysed by Lobato and Savin (1998). Using multiplier test, Lobato and Savin (1998) 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.

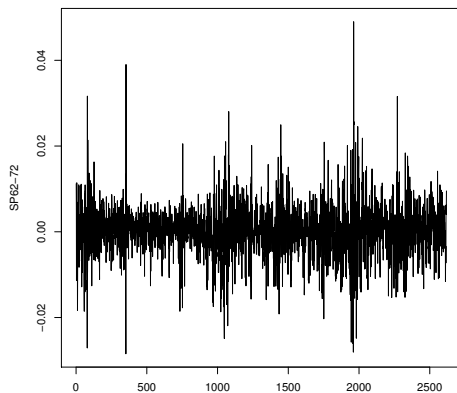


Figure 3: Daily log returns of S&P500, July 1962 to December 1972, 2016 data

In Figure 4(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 the Beta kernel of the log returns and the absolute log returns is proposed in Figures 4(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 the smaller  $b$  is, the 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.

Also, Figures 4(c) and (d) show that the Beta kernel estimator of the spectrum at  $\lambda = 0$  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 4(e) and (f) for the log returns and the absolute log returns, respectively. Three 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 (see Brillinger (2001) for the definition of these kernel functions). Although the three competitors show a peak close to the frequency zero in Figure 4(f), the three kernel methods do not clearly show the expected unboundedness at zero of the spectral density function of the absolute return stock price. 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.



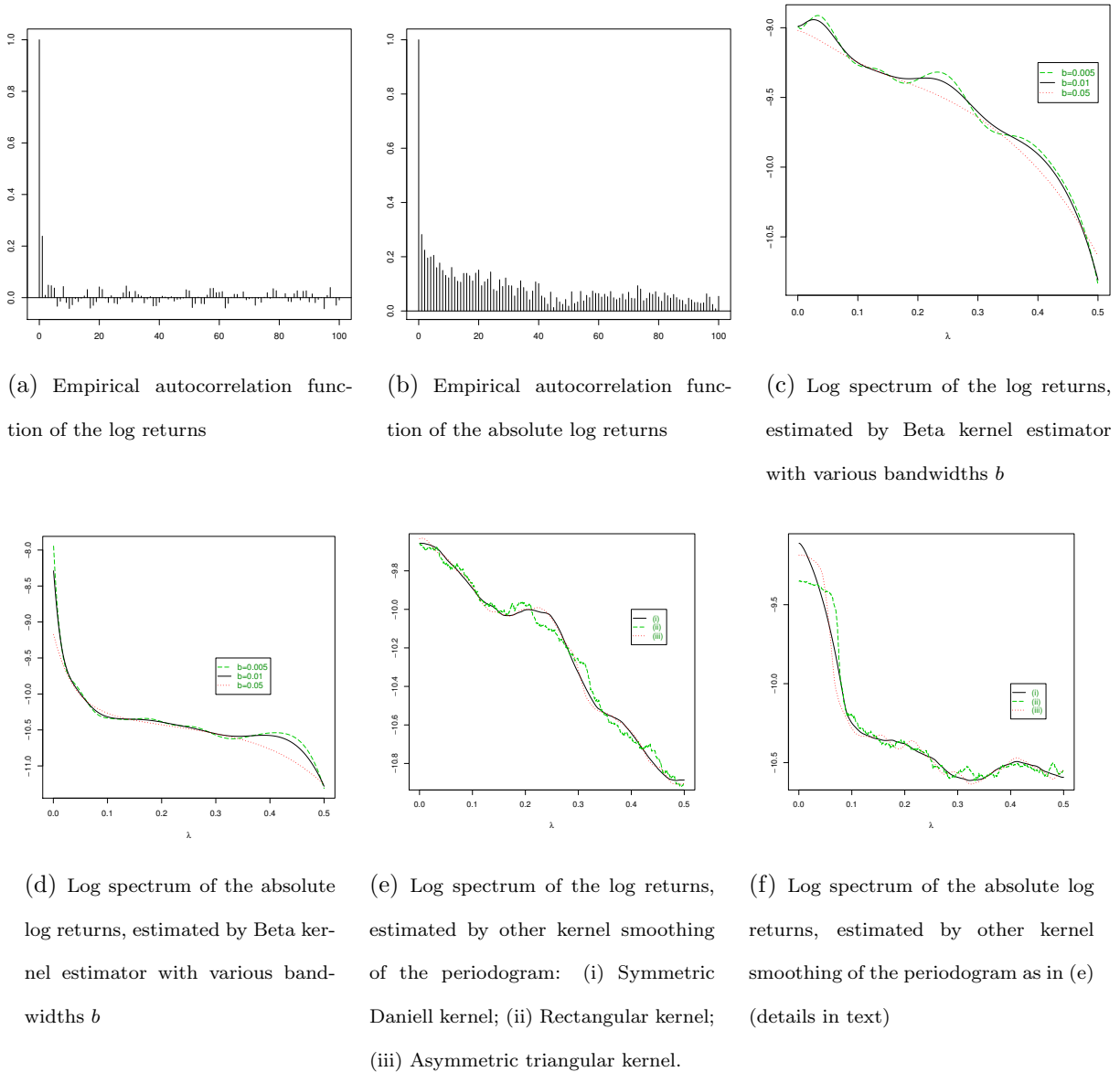


Figure 4: Empirical autocorrelation function and Log spectrum estimation of the daily log returns of S&P500 (July 1962 to December 1972)

## 5 Discussion

The estimation of a spectral density often requires to know whether we observe a long or short memory time series. The periodogram is a natural estimator for the spectral density, however, it is asymptotically unbiased and its variance does not converges to zero. Consequently, the periodogram is not a consistent estimator. Smoothing the periodogram using symmetric

kernels was proposed for short dependence memory time series, see Ferreira and Rodriguez-Poo (1998). But, this method is not consistent for a spectral density that is not null near zero. Furthermore, when we have a long memory time series, the symmetric kernel estimator leads a severe bias at the boundary. The semi-parametric estimator proposed by Robinson (1994a) is frequently used when we estimate the spectral density of long-range dependence time series.

In this paper, we propose a new nonparametric estimator of the spectral density by smoothing the periodogram using Beta kernel. Beta kernel is an asymmetric function with a shape varying according to the frequency where the spectrum is estimated. This estimator doesn't require the knowledge of the nature of the dependence range of the process, and it is adapted for both short and long memory time series. We proved its consistency and showed its finite sample performance. Here, we propose a global bandwidth for the Beta kernel estimator but we suggest to investigate the local bandwidth parameter selection in order to improve the performance of the Beta kernel estimator as proposed for the symmetric kernel estimator.

## A Appendix: Proofs

In this section, we provide the proofs of the theoretical results given in Section 3.

**Proof of Proposition 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

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

By the smoothness conditions assumed on  $f$  it is straightforward to check that the approximation term  $R_1$  has the rate  $O(T^{-1})$ . The term  $R_2$  pools the periodogram over frequencies. After calculation,  $R_2$  can be written as

$$R_2 = \frac{1}{T} \sum_j K_{b,\lambda}(\omega_j) A_{j,T}((0, 1)) \quad \text{with } A_{j,T}(I) := \int_I \mathcal{K}_T(u) \{f(u + \omega_j) - f(\omega_j)\} du.$$

The function  $\mathcal{K}_T(u) := (2\pi T)^{-1} |\sum_s \exp(2\pi ius)|^2$  is the Fejér kernel, and satisfies

$$\mathcal{K}_T(u) = O\left(\frac{T}{1 + T^2 u^2}\right), \text{ for } 0 < |u| \leq 1 \text{ and } \int_0^1 \mathcal{K}_T(u) du = 2\pi,$$

(see Robinson (1994a)). Moreover, the Beta kernel satisfies  $K_{b,\lambda}(u) \leq c_1 b^{-1/2} \{\lambda(1 - \lambda)\}^{-1/2}$ , where  $c_1$  is a positive constant (see Chen (2000)). Let  $m = m(T)$  be a sequence such that  $m = o(T^{1/2} b^{1/4})$ , for which  $\omega_1, \dots, \omega_m$  are sufficiently small.  $R_2$  is equal to the the sum of the three following terms:

$$\begin{aligned} R_{21} &= \frac{1}{T} \sum_{j=1}^m K_{b,\lambda}(\omega_j) A_{j,T}((0, 1)), \\ R_{22} &= \frac{1}{T} \sum_{j=m+1}^T K_{b,\lambda}(\omega_j) A_{j,T}((0, \epsilon]), \\ R_{23} &= \frac{1}{T} \sum_{j=m+1}^T K_{b,\lambda}(\omega_j) A_{j,T}((\epsilon, 1)). \end{aligned}$$

In the expression of  $R_{21}$ , the rate of  $f(\omega_j)$  dominates that of  $f(u + \omega_j)$ , which implies that  $A_{j,T}((0, 1)) = O(f(\omega_j))$  for  $j = 1, \dots, m$ . Hence,

$$R_{21} = O\left(\frac{m}{T\sqrt{b}} \sum_{j=1}^m \frac{\omega_j^{-\beta}}{m}\right) = O\left(\frac{m}{T\sqrt{b}}\right) = o(T^{-1/2} b^{-1/4}).$$

Now, using the Lipschitz property  $|f(u + \omega_j) - f(\omega_j)| \leq Cu$ , for  $u > 0$  and some constant  $C$ , we find that

$$|R_{23}| = O\left(\frac{1}{T} \sum_{j=m+1}^T K_{b,\lambda}(\omega_j) \times \frac{1}{T\epsilon}\right) = O\left(\frac{1}{T\epsilon}\right),$$

and

$$|R_{22}| = O\left(\frac{1}{T} \sum_{j=m+1}^T K_{b,\lambda}(\omega_j) \times \epsilon\right) = O(\epsilon).$$

By choosing  $\epsilon = T^{-1/2}$ , we find that

$$R_2 = o(T^{-1/2}b^{-1/4}).$$

On the other hand, one can write

$$\begin{aligned}\mathbb{E}(\hat{f}(\lambda)) &= \int_0^1 K_{b,\lambda}(u)f(u)du + R_2 \\ &= \mathbb{E}(f(\xi_\lambda)) + R_2\end{aligned}\tag{9}$$

where  $\xi_\lambda$  is a Beta random variable with parameters  $1 + \lambda/b$  and  $1 + (1 - \lambda)/b$ . A general approximation bias of  $\mathbb{E}(f(\xi_\lambda))$  has been derived in Chen (2000) and leads to

$$f(\lambda) - \mathbb{E}(\hat{f}(\lambda)) = b\{(1 - 2\lambda)f'(\lambda) + \frac{1}{2}\lambda(1 - \lambda)f''(\lambda)\} + o(b) + R_2.$$

This ends the calculation of the expectation of the Beta kernel.

Now, the variance of the estimator is

$$\text{Var}\hat{f}(\lambda) = \frac{1}{T^2} \sum_{s,t=1}^T K_{b,\lambda}(\omega_s)K_{b,\lambda}(\omega_t)\text{Cov}(I_T(\omega_s), I_T(\omega_t)).$$

First, for  $t \neq s$  we use the fact that  $\text{Cov}(I_T(\omega_s)/f(\omega_s), I_T(\omega_t)/f(\omega_t)) = r_{st}$  where  $\sum_{s<t} r_{st} = O(\log^r(T))$  for some  $r > 0$  (e.g. Moulines and Soulier (2003)). Recalling that the Beta kernel satisfies  $K_{b,\lambda}(u) \leq c_1 b^{-1/2} \{\lambda(1 - \lambda)\}^{-1/2}$ , the sum over  $s \neq t$  is then of order

$$\frac{1}{bT} \sum_{s<t} r_{st} f(\omega_s) f(\omega_t).\tag{10}$$

Let's examine Equation (10) for  $\omega_s$  close to 0 and away from 0. For  $s$  such that  $\omega_s > \omega_0 > 0$  the rate is  $\log^r(T)/(bT^2)$ . Now, Let  $\{s_T\}$  be a sequence such that  $s_T^{-1} + Ts_T^{-1} \rightarrow 0$ . For  $s$  such that  $\omega_{s_T} > \omega_s$ , the sum in (10) can be decomposed into two terms:

$$\frac{1}{bT} \sum_{s=1}^{s_T} \sum_{t=s+1}^{s_T} r_{st} f(\omega_s) f(\omega_t) + \frac{1}{bT} \sum_{s=1}^{s_T} \sum_{t=s_T}^T r_{st} f(\omega_s) f(\omega_t),$$

where the first term is the dominant term, with rate

$$O\left(\frac{1}{bT^{2-2\beta}} \sum_{s=1}^{s_T} \sum_{t=s+1}^{s_T} s^{-\beta} t^{-\beta} r_{st}\right) = O\left(\frac{\log^r(T)}{bT^{2-2\beta}}\right).$$

Using similar arguments, we can obtain the same rate for  $s$  such that  $\omega_0 > \omega_s > \omega_{sT}$ .

Second, for  $t = s$ , we use the fact that  $\text{Var}(I_T(\omega_s)/f(\omega_s))$  is asymptotically uniformly bounded (e.g. Deo (1997)), so the sum in  $\text{Var}\hat{f}(\lambda)$  is asymptotically equal to

$$\frac{1}{T^2} \sum_s K_{b,\lambda}(\omega_s)^2 f(\omega_s)^2 = \frac{1}{T} G(\lambda) \mathbb{E}(f(\rho_\lambda)^2)$$

where  $\rho_\lambda$  is a Beta random variable with parameters  $1 + 2\lambda/b$  and  $1 + 2(1 - \lambda)/b$  and

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

The final result is derived from a Taylor expansion and the asymptotic properties of  $G(\lambda)$ , which is analogous to Chen (2000) and therefore we skip the detail. This result will be useful for proving Proposition 2.

□

The following lemma establishes the uniform convergence of the bias of the Beta kernel estimator.

**Lemma 1** *If the spectral density  $f$  is a continuous function on the interval  $(0, 1)$ , then for any compact  $I$  in  $(0, 1)$*

$$\sup_{\lambda \in I} \left| \mathbb{E}(\hat{f}(\lambda)) - f(\lambda) \right| \rightarrow 0$$

as  $T \rightarrow \infty$  and  $b = b(T) \rightarrow 0$ .

**Proof** If  $\mu_\lambda$  and  $\sigma_\lambda^2$  are the respective mean and variance of the random variable  $\xi_\lambda$ , where  $\xi_\lambda$  is the Beta random variable with parameters  $1 + \lambda/b$  and  $1 + (1 - \lambda)/b$ , 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  $|\Delta_j(\lambda)| \leq Mb^2$  for  $j = 1$  and  $2$  (see Johnson et al. (2000)).

To prove the lemma, we first consider the approximation  $\mathbb{E}(\hat{f}(\lambda))$  by  $\int_0^1 K_{b,\lambda}(u)f(u)du$

given in (9). Consider the following decomposition of the dominant term in (9):

$$\begin{aligned}
& \left| \int_0^1 \{f(t) - f(\lambda)\} K\{t, \lambda/b + 1, (1 - \lambda)/b + 1\} dt \right| \\
& \leq 2 \int_{|t - \mu_\lambda| < \delta} |f(t) - f(\lambda)| K\{t, \lambda/b + 1, (1 - \lambda)/b + 1\} dt + 2 \int_{\mu_\lambda + \delta}^1 (\dots) + 2 \int_0^{\mu_\lambda - \delta} (\dots) \\
& = \text{I} + \text{II} + \text{III}.
\end{aligned}$$

We want to show the convergence of I, II and III to zero.

First, 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  $\text{I} \leq \epsilon$  for all  $b \leq b_\epsilon^{\text{I}}$ .

Second, since  $\sigma_\lambda^2 \leq (b + 4Mb^2)$ , then by Chebyshev's inequality,

$$\begin{aligned}
\text{II} & \leq 4 \Pr(Z - \mu_\lambda > \delta) \sup_{t > \mu_\lambda + \delta} |f(t)| \\
& \leq \frac{4}{\delta^2} \sigma_\lambda^2 \sup_{t > \mu_\lambda + \delta} |f(t)| \\
& \leq \frac{1}{\delta^2} (b + 4Mb^2) \sup_{t > \mu_\lambda + \delta} |f(t)| \\
& \leq \epsilon \quad \text{for all } b \leq b_\epsilon^{\text{II}}.
\end{aligned}$$

Finally, 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  $\xi_{\lambda, \beta}$  denotes the Beta random variable with parameters  $1 + \lambda/b - \beta$  and  $1 + (1 - \lambda)/b$ , then  $\mathbb{V}\text{ar}(\xi_{\lambda, \beta}) = b\lambda(1 - \lambda) + O(b^2)$ . Therefore,

$$\begin{aligned}
\text{III} & \leq 4 \int_0^{\mu_\lambda - \delta} t^{-\beta} K\{t, \lambda/b + 1, (1 - \lambda)/b + 1\} dt \\
& = \frac{4B\{\lambda/b - \beta + 1, (1 - \lambda)/b + 1\}}{B\{\lambda/b + 1, (1 - \lambda)/b + 1\}} \int_0^{\mu_\lambda - \delta} K\{t, \lambda/b - \beta + 1, (1 - \lambda)/b + 1\} dt \\
& \leq \frac{4B\{\lambda/b - \beta + 1, (1 - \lambda)/b + 1\}}{B\{\lambda/b + 1, (1 - \lambda)/b + 1\}} \frac{\mathbb{V}\text{ar}(\xi)}{\delta^2} \\
& \leq \epsilon \quad \text{for all } b \leq b_\epsilon^{\text{III}}.
\end{aligned}$$

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

□

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

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

for some constant  $C$  depending on  $\eta_1$  and  $\eta_2$ . Therefore, for  $\lambda$  and  $\lambda' \in I$ , we have

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

Hence, using similars arguments as for the smaller order approximation terms of the expectation in Proposition 1, we obtain that

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

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

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

Therefore,

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

Using Proposition 1 and Chebychev's inequality, we can show that  $\hat{f}(\lambda) - \mathbb{E}\hat{f}(\lambda) = O_P(b^{-1/2}T^{-1})$  for all  $\lambda \in I$ . Therefore  $\max_j |\hat{f}(\lambda_j) - \mathbb{E}\hat{f}(\lambda_j)| = O_P(Nb^{-1/2}T^{-1}) = O_P(b^{-2-\epsilon}T^{-1})$  which concludes the proof.  $\square$

### Proof of Proposition 3.

Proof of (i): 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 Equation (9) 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 in the proof of Proposition 1 on the pooled periodogram in order to show that  $R_2 = o(1)$  under the constraints on  $b$  stated in the proposition. Therefore, for  $\delta$  sufficiently small,

$$\begin{aligned} \mathbb{E}(\hat{f}(0)) &= b^{-1}(1+b) \int_0^1 (1-u)^{1/b} f(u)du + o(1) \\ &> b^{-1}(1+b)C \int_0^\delta (1-u)^{1/b} du + o(1) \\ &> C(1 - (1-\delta)^{1/b+1}) + o(1) \\ &\rightarrow C, \quad \text{as } b \rightarrow 0. \end{aligned}$$

Hence, the divergence of  $\mathbb{E}(\hat{f}(0))$  is proved. To show the convergence in probability, we use the Chebychev's inequality and Proposition 1. Indeed, for any  $\epsilon > 0$  and for a sequence  $\lambda$  such that  $\lambda/b \rightarrow \kappa$ ,

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

Now using the fact that  $K_{b,\lambda}(t)$  tends to  $K_{b,0}(t)$ , as  $\lambda$  converges 0, we obtain

$$\left|\int_0^1 K_{b,\lambda}(u)f(u)du - \int_0^1 K_{b,0}(u)f(u)du\right| = o(1), \text{ as } \lambda \rightarrow 0.$$



Hence,  $\mathbb{E}\hat{f}(\lambda) \rightarrow \mathbb{E}\hat{f}(0)$ , as  $\lambda \rightarrow 0$ . Which concludes the proof of (i).

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

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

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  $t < \delta$ . Calculating the integral over  $[0, \delta]$  and  $[\delta, 1]$ , 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  $b$  converges 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, we conclude with Chebychev inequality and Proposition 1 that lead to  $P(|\hat{f}(0) - \mathbb{E}\hat{f}(0)| > \epsilon) = O(b^{-1}T^{-1})$  and obtain the stated result. □

**Proof of Proposition 4.** We start by proving the relative convergence of the bias term, that is,  $\left| \frac{\mathbb{E}(\hat{f}(\lambda)) - f(\lambda)}{f(\lambda)} \right|$  goes to zero as  $\lambda/b$  tends to  $\kappa$ . We proceed as in the beginning of the proof of Lemma 1. Omitting the negligible terms, we use the decomposition

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

in which  $\text{I} \leq \epsilon$  for all  $b \leq b_\epsilon^{\text{I}}$  and  $\text{II} \leq \epsilon$  for all  $b \leq b_\epsilon^{\text{II}}$ . The treatment of the term III is not as in Lemma 1. Using Taylor expansion around  $\lambda$  and that  $f'(\lambda) = O(\lambda^{-\beta-1})$  for small value of  $\lambda$ , we obtain

$$\begin{aligned} \text{III} &\simeq \frac{|f'(\lambda)|}{f(\lambda)} \int_0^{\mu_\lambda - \delta_\epsilon} (\lambda - t) K\{t, \lambda/b + 1, (1-\lambda)/b + 1\} dt \\ &\leq \frac{|\lambda f'(\lambda)|}{f(\lambda)} \frac{\text{Var}(\xi_\lambda)}{\delta_\epsilon^2} \quad (\text{Chebyshev inequality}) \\ &\leq \epsilon \end{aligned}$$

for all  $b \leq b_\epsilon^{\text{III}}$ , where  $\xi_{\lambda,\beta}$  is a Beta random variable with parameters  $1 + \lambda/b - \beta$  and  $1 + (1 - \lambda)/b$ , with  $\text{Var}(\xi) = b\lambda(1 - \lambda) + O(b^2)$ . By combining the three terms, the bias term is bounded by  $3\epsilon$  for all  $b \leq \min(b_\epsilon^{\text{I}}, b_\epsilon^{\text{II}}, b_\epsilon^{\text{III}})$ .

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

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

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

## References

- M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover, 1972.
- D. W. K. Andrews and Y. Sun. Adaptive local polynomial whittle estimation of long-range dependence. *Econometrica*, 72:569–614, 2004.
- S. Birr, S. Volgushev, T. Kley, H. Dette, and M. Hallin. Quantile spectral analysis for locally stationary time series. *J. R. Statist. Soc. B*, 79:1619–1643, 2017.
- D. R. Brillinger. *Time series: Data analysis and theory*. SIAM, 2001.
- B. M. Brown and S. X. Chen. Beta-bernstein smoothing for regression curves with compact supports. *Scand. J. Statist.*, 26:47–59, 1999.
- S. X. Chen. Beta kernel estimators for density functions. *Comput. Statist. Data. Anal.*, 31: 131–145, 1999.
- S. X. Chen. Beta kernel smoothers for regression curves. *Statist. Sinica*, 10:73–91, 2000.

- M. Delgado and P. M. Robinson. Optimal spectral bandwidth for long memory. *Statist. Sinica*, 6:97–112, 1996a.
- M. Delgado and P. M. Robinson. Optimal spectral kernel for long-range dependent time series. *Statistics and Probability Letters*, 30:37–43, 1996b.
- R. S. Deo. Asymptotic theory for certain regression models with long memory errors. *Journal of Time Series Analysis*, 18:385–393, 1997.
- P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors. *Theory of Long Range Dependence*. Birkhäuser, Boston, 2003.
- J. Fan and E. Kreutzberger. Automatic local smoothing for spectral density estimation. *Scand. J. Statist.*, 25:359–369, 1998.
- E. Ferreira and J.M. Rodriguez-Poo. Variable bandwidth kernel estimators of the spectral density. *Journal of Time Series Analysis*, 20 (3):271–287, 1998.
- John A. Gubner. Theorems and fallacies in the theory of long-range-dependent processes. *IEEE Trans. Inform. Theory*, 51(3):1234–1239, 2005. ISSN 0018-9448. doi: 10.1109/TIT.2004.842768. URL <https://doi.org/10.1109/TIT.2004.842768>.
- J. Hannig and T.C.M. Lee. Kernel smoothing of periodograms under kullback-leibler discrepancy. *Signal Processing*, 84:1255–1266, 2004.
- M. Henry. Bandwidth choice, optimal rates and adaptivity in semiparametric estimation of long memory. In Gilles Teyssière and Alan P. Kirman, editors, *Long Memory in Economics*, pages 157–172. Springer, 2007.
- Yaozhong Hu, David Nualart, Samy Tindel, and Fangjun Xu. Density convergence in the breuer-major theorem for gaussian stationary sequences. *Bernoulli*, 21(4):2336–2350, 11 2015. doi: 10.3150/14-BEJ646. URL <https://doi.org/10.3150/14-BEJ646>.
- C. M. Hurvich. Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods. *J. Amer. Statist. Assoc.*, 14:227–238, 1980.

- N. L. Johnson, Samuel Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*. Wiley-Interscience, 2nd edition, 2000.
- T.C.M. Lee. A simple span selector for periodogram smoothing. *Biometrika*, 84:965–969, 1997.
- T.C.M. Lee. A stabilized bandwidth selection method for kernel smoothing of the periodogram. *Signal Processing*, 81:419–430, 2001.
- I. N. Lobato and P. M. Robinson. A nonparametric test for  $I(0)$ . *Review of Economic Studies*, 65:475–495, 1998.
- I. N. Lobato and N. E. Savin. Real and spurious long memory properties of stock market data. *J. Bus. Econ. Statist.*, 16:261–283, 1998.
- I. N. Lobato and C. Velasco. Long memory in stock-market trading volume. *J. Bus. Econ. Statist.*, 18:410–427, 2000.
- É. Moulines and P. Soulier. Semiparametric spectral estimation for fractional processes. In Doukhan et al. (2003), pages 251–302.
- A. Ohanissian, J. R. Russel, and R. A. Tsay. True or spurious long memory? A new test. *J. Bus. Econ. Statist.*, 26:161–175, 2008.
- W. Palma. *Long-Memory Time Series: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley, 2007.
- K.S. Riedel and A. Sidorenko. Adaptive smoothing of the log-spectrum with multiple tapering. *IEEE Trans. Signal Process*, 44:1794–800, 1996.
- P. M. Robinson. Semiparametric analysis of long-memory time series. *Ann. Statist.*, 22: 515–539, 1994a.
- P. M. Robinson. Rates of convergence and optimal spectral bandwidth for long range dependence. *Probab. Theory Related Fields*, 99:443–473, 1994b.

P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23:1630–1661, 1995.

P. M. Robinson, editor. *Time Series with Long Memory*. Oxford University Press, 2003.

P. M. Robinson and M. Henry. Higher-order kernel semiparametric  $M$ -estimation of long memory. *J. Econometrics*, 114:1–27, 2003.