Adaptive Bayesian Estimation of a spectral density

Judith Rousseau[a,b], Kruijer Willem[c]

[a] *CEREMADE*
Université Paris Dauphine,
Place du Maréchal de Lattre de Tassigny
75016 Paris, FRANCE.
[b] *ENSAE-CREST,*
3 avenue Pierre Larousse,
92245 Malakoff Cedex, FRANCE .
[c] *Wageningen University*
Droevendaalsesteeg 1
Biometris, Building 107
6708 PB Wageningen
The Netherlands

Abstract

Rousseau et al. [8] recently studied the asymptotic behavior of Bayesian estimators in the FEXP-model for spectral densities of Gaussian time-series. For the L_2-norm on the log-spectral densities, they proved that the convergence rate is at least $n^{-\frac{\beta}{2\beta+1}}(\log n)^{\frac{2\beta+2}{2\beta+1}}$, $\beta > \frac{1}{2}$ being the Sobolev-regularity of the true spectral density f_o. We will improve upon the logarithmic factor, and prove that given a prior only depending on $\beta_s > \frac{1}{2}$, we have adaptivity to any $\beta \geq \beta_s$.

Keywords: Bayesian non-parametric, rates of convergence, adaptive estimation, long-memory time-series, FEXP-model

1. Introduction

Let X_t, $t \in \mathbb{Z}$, be a stationary zero mean Gaussian time series with spectral density $f_o(\lambda)$, $\lambda \in [-\pi, \pi]$ in the form

$$f_o(\lambda) = |1 - e^{i\lambda}|^{-2d_o} \exp\left\{\sum_{j=0}^{\infty} \theta_{o,j} \cos(j\lambda)\right\}, \quad \theta_o \in \Theta(\beta, L_o) \tag{1.1}$$

where $d_o \in (-\frac{1}{2}, \frac{1}{2})$, $\Theta(\beta, L_o) = \{\theta \in l_2(\mathbb{N}) : \sum_{j\geq 0} \theta_j^2(1+j)^{2\beta} \leq L_o\}$ is a Sobolev ball. The parameter d_o is called the long-memory parameter; we will refer to $\exp\{\sum_{j=0}^{\infty} \theta_{o,j} \cos(j\lambda)\}$ as the short-memory part of the spectral density. The parameter β controls the regularity of the short-memory part. It is then natural to use the fractionally exponential or FEXP-model (see Beran [2] and Moulines and Soulier [6] and references therein) $\mathcal{F} = \cup_{k\geq 0}\mathcal{F}_k$, where

$$\mathcal{F}_k = \left\{ f_{d,k,\theta}(\lambda) = |1-e^{i\lambda}|^{-2d} \exp\left\{\sum_{j=0}^{k} \theta_j \cos(j\lambda)\right\},\ d \in \Big(-\frac{1}{2}, \frac{1}{2}\Big), \theta \in \mathbb{R}^{k+1} \right\}.$$

We study Bayesian estimation of f_o within this FEXP-model. Let $\pi(d,k,\theta)$ denote the prior on (d,k,θ); this induces a prior on $\mathcal{F}$ which we also denote π. Let $T_n(f)$ denote the covariance matrix of the observations $X = (X_1,\dots,X_n)$, and let l_n be the associated log-likelihood

$$l_n(d,k,\theta) = -\frac{k+1}{2}\log(2\pi) - \frac{1}{2}\log|T_n(f)| - \frac{1}{2}X' T_n^{-1}(f) X \tag{1.2}$$

Bayesian estimates of the spectral density f_o are based on the posterior

$$\pi(f \in A|X) = \frac{\int_A e^{l_n(d,k,\theta)} d\pi(f)}{\int_{\mathcal{F}} e^{l_n(d,k,\theta)} d\pi(f)}, \qquad A \subset \mathcal{F}. \tag{1.3}$$

For example the posterior mean or median could be taken as 'point'-estimators of f_o. In this work however we focus on the posterior itself, and study the rate of convergence at which the posterior concentrates at f_o. More precisely, we lower-bound the posterior mass on the sets

$$B(\epsilon_n) = \{f \in \mathcal{F} : l(f, f_o) \leq \epsilon_n^2\},$$

where ϵ_n is a sequence tending to zero and

$$l(f, f_o) = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\log f_o(\lambda) - \log f(\lambda))^2 d\lambda.$$

Whether $\pi(B(\epsilon_n)|X)$ tends to one for a certain sequence ϵ_n critically depends on the smoothness of f_o as well as the smoothness induced by the prior. In Theorem 4.2 of Rousseau et al. [8] (RCL hereafter) it is shown that when $\theta_o \in \Theta(\beta, L_o)$ and the prior on θ has support contained in a Sobolev ball $\Theta(\beta, L)$ with L large enough, then the rate is $\epsilon_0(L) n^{-\frac{2\beta}{2\beta+1}} (\log)^{\frac{4\beta+4}{2\beta+1}}$, for fixed $\beta > \frac{1}{2}$ and $\epsilon_0(L)$ large enough depending on L. In the present work we prove that such priors in fact lead to an adaptive concentration rate (in β) and we improve upon the constant $\epsilon_0(L)$ and the logarithmic factor. Adaptivity is of great interest since it is difficult to know the smoothness of the function f a priori. Improving on the constant ϵ_0 is crucial in Kruijer and Rousseau [4] but has also interest in its own. Indeed in Theorem 2.1 we prove that ϵ_0 depends only on L_o the radius of the Sobolev ball containing θ_o. In RCL however ϵ_0 depends on L, with the risk that would L be very large ϵ_0 might also be very large. Here we prove that this is not the case and that we can choose L as large as the application requires. This suggests that the result might actually hold without the constraint L in the prior on θ, but we have not been able to prove that.

Notation:
The m-dimensional identity matrix is denoted I_m. For matrices A we write $|A|$ for the Frobenius or Hilbert-Schmidt norm $|A| = \sqrt{\mathrm{tr} AA^t}$, where A^t denotes the transpose of A. The operator or spectral norm is denoted $\|A\|^2 = \sup_{\|x\|=1} x'Ax$. We also use $\|\cdot\|$ for the Euclidean norm of finite dimensional vectors or sequences in $l^2(\mathbb{N})$, and for the L_2-norm of functions. If $u \in l^1(\mathbb{N})$ we denote $\|u\|_1 = \sum_j |u_j|$. Given a sequence $\{u_j\}_{j\geq 0}$ and a nonnegative integer m, we write $u_{[m]}$ for the vector $(u_0, \dots, u_m)$ and $\|u\|_{>m}$ for the l^2-norm of the sequence $u_{m+1}, u_{m+2}, \dots$. When we write $\sum_{j\geq 0}(\theta_j - \theta_{o,j})^2$ or $\sum_{j\geq 0}|\theta_j - \theta_{o,j}|$ for a finite-dimensional vector θ and $\theta_o \in l_2(\mathbb{N})$, θ_j is understood to be zero when $j > k$. For any function $h \in L_1([-\pi, \pi])$, $T_n(h)$ is the matrix with entries $\int_{-\pi}^{\pi} e^{i|l-m|\lambda} h(\lambda) d\lambda$, $l, m = 1, \dots, n$. For example, $T_n(f)$ is the covariance matrix of observations $X = (X_1, \dots, X_n)$ from a time series with spectral density f. Let P_o denote the law associated with the true spectral density f_o and E_o expectations with respect to P_o.

2. Main results

Let $\beta_s > \frac{1}{2}$ be a fixed constant. We consider the following family of priors on (d, k, θ). d is a priori independent of (k, θ) with density π_d with respect to Lebesgue measure. For some positive $t < 1/2$, the support of π_d is included in $[-1/2 + t, 1/2 - t]$. We consider two cases for the prior on k:

Deterministic sieve $\pi_k(k) = \delta_{k_{A,n}}(k)$, i.e. it is the Dirac mass at $k_{A,n} = \lfloor A(n/\log n)^{1/(2\beta_s+1)} \rfloor$, for some positive A.

Random sieve the support of π_k is $\mathbb{N}$ and satisfies:

$$e^{-c_1 k \log k} \leq \pi_k(k) \leq e^{-c_2 k \log k},$$

for some positive c_1, c_2 and k large enough. $\pi_{\theta|k}$, the prior on θ given k, has a density with respect to the Lebesgue measure on $\mathbb{R}^k$. This density is also denoted $\pi_{\theta|k}$, and is such that, for some constants $L > 0$ and $\beta_s > 1/2$, $\pi_{\theta|k}$ is positive on $\Theta_k(\beta, L)$ and $\pi_{\theta|k}[\Theta_k(\beta_s, L)^c] = 0$. These priors have been considered in particular in RCL, in Holan et al. [3] and in Kruijer and Rousseau [4]. We now state the main result.

Theorem 2.1. *Suppose we observe* $X = (X_1, \ldots, X_n)$ *from a stationary, zero mean Gaussian time-series whose spectral density* f_o *is as in* (1.1), *with* $d_o \in [-\frac{1}{2} + t, \frac{1}{2} - t]$, $\theta_o \in \Theta(\beta, L_o)$ *and* $\beta \geq \beta_s > \frac{1}{2}$. *Consider a prior* $\pi = \pi_d \pi_k \pi_{\theta|k}$ *as described above such that there exists* $c_0 > 0$ *for which*

$$\liminf_{n\to\infty} \min_{k \in \mathcal{K}_n} \inf_{\theta \in \Theta_k(\beta, L_o)} e^{c_0 k \log k} \pi_{\theta|k}(\theta) > 1. \tag{2.1}$$

where, for some $B > 0$ *and* $k_{B,n} = \lfloor B(n/\log n)^{1/(2\beta_s+1)} \rfloor$, $\mathcal{K}_n = \{0, \ldots, k_{B,n}\}$ *in the case of the random sieve prior, and* $\mathcal{K}_n = \{k_{A,n}\}$ *in the case of the deterministic prior. Assume also that* L *is large enough.*

- *In the case of the random sieve prior, for any* $\beta_2 > \beta_s$, *we have the following uniform result:*

$$\sup_{f_o \in \cup_{\beta_s \leq \beta \leq \beta_2} \Theta(\beta, L_o)} E_o \pi((d, k, \theta) : l(f_{d,k,\theta}, f_o) \geq l_0^2 \epsilon_n^2(\beta) | X) \leq n^{-3}, \tag{2.2}$$

  where $\epsilon_n(\beta) = (n/\log n)^{-\frac{\beta}{2\beta+1}}$ *and* l_0 *only depends on* L_o. *In particular, it is independent of* L.

- *In the case of the deterministic prior, for any* $\beta_2 > \beta_s$, *we have the following uniform result:*

$$\sup_{f_o \in \cup_{\beta_s \leq \beta \leq \beta_2} \Theta(\beta, L_o)} E_o \pi((d, k, \theta) : l(f_{d,k,\theta}, f_o) \geq l_0^2 \epsilon_n^2(\beta_s) | X) \leq n^{-3}, \tag{2.3}$$

  where l_0 *only depends on* L_o *and is independent of* L.

The constraint $\beta > 1/2$ is necessary to ensure that the short memory part $\exp(\sum_j \theta_{oj} \cos(jx))$ is bounded and continuous. As mentioned in the introduction, the fact that l_0 is independent of L is interesting since it allows us, in practice, to choose L arbitrarily high without penalizing the posterior concentration rate. It suggests that such results could hold with $L = \infty$, however we have no proof for it. The random sieve prior leads to an adaptive posterior concentration rate over the range $\beta \geq \beta_s$, since for all $\beta > 1/2$, $\epsilon_n(\beta)$ is the minimax (up to a $\log n$ term) rate over the class of FEXP spectral densities given by (1.1) and associated to $\theta \in \Theta(\beta, L_o)$. The deterministic sieve prior does not lead to an adaptive procedure since the posterior concentration rate is $\epsilon_n(\beta_s)$ in this case. Obtaining adaptation by putting a prior on the dimension of the model is a commonly used strategy in Bayesian non parametrics, see for instance Arbel [1] or Rivoirard and Rousseau [7].

3. Proof of Theorem 2.1

We first introduce some notions that are useful throughout the proof.

3.1. Notation and preliminary results

We first introduce various (pseudo)-distances. We denote the Kullback - Leibler divergence between the Gaussian distributions associated with spectral densities f_o and f by

$$KL_n(f_o; f) = \frac{1}{2n}\left\{\operatorname{tr}\left[T_n(f_o)T_n^{-1}(f) - I_n\right] - \log\det(T_n(f_o)T_n^{-1}(f))\right\},$$

a symmetrized version of it by $h_n(f_o, f) = KL_n(f_o; f) + KL_n(f; f_o)$ and the variance of the log-likelihood ratio by

$$b_n(f_o, f) = \frac{1}{n}\operatorname{tr}\left\{T_n^{-1}(f)(T_n(f_o - f)T_n^{-1}(f)T_n(f_o - f)\right\}.$$

The limiting values of $b_n(f_o, f)$ and $h_n(f_o, f)$ are denoted

$$h(f_o, f) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\frac{f_o(\lambda)}{f(\lambda)} + \frac{f(\lambda)}{f_o(\lambda)} - 2\right]d\lambda, \quad b(f_o, f) = (2\pi)^{-1}\int_{-\pi}^{\pi}\left(\frac{f_o}{f}(\lambda) - 1\right)^2 d\lambda.$$

Then $h(f_o, f) \geq l(f_o, f)$ (RCL, p.6). Using Lemma 2 in RCL we find that for all $k \in \mathbb{N}$,

$$\begin{aligned} b_n(f_o, f_{d,k,\theta}) &\leq \|T_n(f_o)^{1/2}T_n(f)^{-1/2}\|^2 h_n(f_o, f_{d,k,\theta}) \\ &\leq C(\|\theta_o\|_1 + \|\theta\|_1)n^{2(d_o - d)_+}h_n(f_o, f), \end{aligned} \tag{3.1}$$

where C is a universal constant. Similarly,

$$h_n(f_o, f) \leq \|T_n^{-\frac{1}{2}}(f)T_n^{\frac{1}{2}}(f_o)\|^2 b_n(f_o, f). \tag{3.2}$$

In line with the notation of (1.2), let $\phi(x; d, k, \theta)$ denote the density of X, which is the Gaussian density with mean zero and covariance matrix $T_n(f_{d,k,\theta})$ and let $\phi(x; d_o, \theta_o)$ denote the Gaussian density associated with $T_n(f_o)$. We write $R_n(f_{d,k,\theta}) = \phi(X; d, k, \theta)/\phi(X; d_o, \theta_o)$ for the likelihood-ratio.

The proof of Theorem 2.1 contains two parts. First, it needs to be shown that the rate is $l_0^2\epsilon_n^2$, for a constant l_0 that may depend on L and β_s. Then by re-insertion of the rate obtained in the first part, we improve upon the constant l_0. In particular, it is shown to be independent of L for L large enough.

3.2. Proof of Theorem 2.1

Throughout the proof C denotes a universal constant. Let $0 < t < 1/2$ and

$$\mathcal{G}_k(t, \beta_s, L) = \left\{ f_{d,k,\theta} : d \in [-\frac{1}{2} + t, \frac{1}{2} - t], \theta \in \Theta_k(\beta_s, L) \right\}, \quad \mathcal{G} = \cup_{k=0}^{\infty} \mathcal{G}_k(t, \beta_s, L).$$

By the results of RCL (Theorem 3.1, and Corollary 1 in the supplement) we have consistency for $h(f_o, f_{d,k,\theta})$ and $|d - d_o|$, i.e. for all $\delta, \epsilon > 0$, $\pi\left(f_{d,k,\theta} : h(f_{d,k,\theta}, f_o) < \epsilon^2, |d - d_o| < \delta | X\right)$ tends to one in probability. Hence it suffices to show that

$$\pi\left[W_n | X\right] = \frac{\int_{W_n} R_n(f) d\pi(f)}{\int R_n(f) d\pi(f)} := \frac{N_n}{D_n} \xrightarrow{P_o} 0, \tag{3.3}$$

where in the case of the random sieve prior,

$$W_n = \left\{ f_{d,k,\theta} \in \mathcal{G} : l(f_o, f_{d,k,\theta}) \geq l_0 \epsilon_n^2(\beta), h(f_o, f_{d,k,\theta}) \leq \epsilon^2,\ |d - d_o| \leq \delta \right\},$$

for a constant $l_0 > 0$ depending only on L_o, β_s and the prior on k. In the case of the deterministic sieve prior, we replace $\epsilon_n^2(\beta)$ in this definition by $\epsilon_n^2(\beta_s)$. We present the proof of (3.3) for the case of the random sieve prior; the proof for the case of the deterministic sieve prior can be deduced by replacing β by β_s. The proof consists of two parts: first we show that for some $c > 0$,

$$P_o\left[D_n < e^{-2nu_0\epsilon_n^2(\beta)}/2 \right] \leq e^{-cn\epsilon_n^2(\beta)}, \tag{3.4}$$

for which we will establish a lower bound the prior mass on a Kullback-Leibler neighborhood of f_o. In the second part we show that under the event

$D_n \geq \frac{1}{2}e^{-nu_0\epsilon_n^2(\beta)}$ we can control N_n/D_n. This will be done by giving a bound on the upper-bracketing entropy of the model.

For the proof of (3.4), note that RCL already found that if $\beta \geq \beta_s > 1/2$, there exists $u_0 \geq 0$ depending only on L_o such that

$$P_o\left[D_n < e^{-nu_0\epsilon_n^2(\beta)(\log n)^{1/(2\beta+1)}}/2\right] = o(n^{-1}).$$

To prove (3.4), we thus need to improve on the $\log n$ term in the preceding equation. Set

$$\bar{\mathcal{B}}_n = \{(d,k,\theta); KL_n(f_o, f_{d,k,\theta}) \leq \frac{\epsilon_n^2(\beta)}{4}, b_n(f_o, f_{d,k,\theta}) \leq \epsilon_n^2(\beta), d_o \leq d \leq d_o+\delta\},$$

for some positive δ. Recall that

$$\begin{aligned} D_n &= \sum_k \pi_k(k)\int e^{l_n(d,k,\theta)-l_n(f_o)}d\pi_{\theta|k}(\theta)d\pi_d(d), \\ P_o^n\left[D_n < e^{-2nu_0\epsilon_n^2(\beta)}\right] &\leq P_o^n\left[\int_{\bar{\mathcal{B}}_n} e^{l_n(f)-l_n(f_o)}d\pi(f) < e^{-2nu_0\epsilon_n^2(\beta)}\right]. \end{aligned}$$

From the proof of Theorem 4.1 in RCL (section 5.2.1), it follows that

$$P_0^n\left(D_n \leq e^{-nu_0\epsilon_n^2(\beta)}\pi(\bar{\mathcal{B}}_n)/2\right) \leq e^{-Cn\epsilon_n^2(\beta)}$$

for some constant $C > 0$ (independent of L). We now show that

$$\pi(\bar{\mathcal{B}}_n) \geq e^{-nu_0\epsilon_n^2(\beta)/4}. \tag{3.5}$$

Define

$$\tilde{\mathcal{B}}_n = \{(d, k_{B,n}, \theta); d_o \leq d \leq d_o+\epsilon_n^2(\beta)n^{-a}, |\theta_j-\theta_{o,j}| \leq (1+j)^{-\beta}\epsilon_n(\beta)n^{-a}, j = 0,\dots,k_{B,n}\}.$$

We first prove that $\pi(\tilde{\mathcal{B}}_n) \geq e^{-nu_0\epsilon_n^2(\beta)/4}$ and then that $\tilde{\mathcal{B}}_n \subset \bar{\mathcal{B}}_n$. As in RCL (see the paragraph following equation (29) on p. 26), we find that

$$\sum_{j=1}^{\infty}(1+j)^{2\beta}\theta_j^2 \leq 2L_o + \epsilon_n^2(\beta)n^{-2a} \leq 3L_o, \forall\theta \in \tilde{\mathcal{B}}_n$$

for n large enough. Combined with condition (2.1) on $\pi_{\theta|k}$, this implies

$$\pi(\tilde{\mathcal{B}}_n) \geq \left(c\epsilon_n(\beta)k_{B,n}^{-\beta}n^{-a}\right)^{k+3} e^{-c_0k_{B,n}\log k_{B,n}} \geq e^{-c(\beta_s)k_0n\epsilon_n^2(\beta)}, \quad \forall\beta \geq \beta_s.$$

This achieves the proof of (3.4) with $u_0 = c(\beta_s)k_0$.

To show that $\tilde{\mathcal{B}}_n$ is included in $\bar{\mathcal{B}}_n$, first note that equation (3.1) implies that it is enough to bound $h_n(f_o, f)$ on $\tilde{\mathcal{B}}_n$. To this end, we use the decomposition $f_o = f_{o,k_{B,n}} e^{\Delta_{d_o,k_{B,n}}}$, where $f_{o,k_{B,n}} = f_{d_o,k_{B,n},\theta_o}$ and

$$\Delta_{d_o,k_{B,n}}(\lambda) = \sum_{j=k_{B,n}+1}^{\infty} \theta_{o,j}\cos(jx), \quad \forall \lambda \in [-\pi,\pi].$$

Then we have the expansion

$$f_o = f_{o,k_{B,n}}(1 + \Delta_{d_o,k_{B,n}} + \Delta^2_{d_o,k_{B,n}}/2 + O(\Delta^3_{d_o,k_{B,n}})), \quad |\Delta_{d_o,k_{B,n}}|_\infty = o(1)$$

and

$$h_n(f_o, f) \le 2[h_n(f_o, f_{o,k_{B,n}}) + h_n(f_{o,k_{B,n}}, f)]. \tag{3.6}$$

We first deal with the first term above. Let $b_{o,n} = e^{\Delta_{d_o,k_{B,n}}} - 1$ and without loss of generality we can assume that $b_{o,n}$ is positive in the expression of $h_n(f_o, f_{o,k_{B,n}})$ so that for all $\beta > 1/2$,

$$\begin{aligned} h_n(f_o, f_{o,k_{B,n}}) &:= \frac{1}{2n} tr\left[T_n^{-1}(f_o)T_n(f_o b_{o,n})T_n^{-1}(f_o)T_n(f_o b_{o,n})\right] \\ &\le \frac{c}{n} tr\left[T_n^{-1}(g_o)T_n(g_o b_{o,n})T_n^{-1}(g_o)T_n(g_o b_{o,n})\right] \\ &= \frac{c}{n} tr[T_n^2(b_{o,n})] + c\gamma_1 + c\gamma_2 \end{aligned} \tag{3.7}$$

where $g_o(\lambda) = |\lambda|^{-2d_o}$ and c depends only on $\sum_{j=0}^{\infty} |\theta_{o,j}| \le L_o^{1/2}(2\beta - 1)^{-1/2}$, and

$$\begin{aligned} \gamma_1 &= \frac{1}{n}\left(tr\left[\left(T_n(\frac{1}{4\pi^2 g_o})T_n(g_o b_{o,n})\right)^2\right] - tr[T_n^2(b_{o,n})]\right) \\ \gamma_2 &= \frac{1}{n}\left(tr\left[(T_n^{-1}(g_o)T_n(g_o b_{o,n}))^2\right] - tr\left[\left(T_n(\frac{1}{4\pi^2 g_o})T_n(g_o b_{o,n})\right)^2\right]\right). \end{aligned}$$

We first bound the first term of the right hand side of (3.7). Note that $b_{o,n}(\lambda) = \tilde{\Delta}_{d_o,k_{B,n},K_n} + R_0$, where

$$\tilde{\Delta}_{d_o,k_{B,n},K_n}(\lambda) = \sum_{j=k_{B,n}+1}^{K_n} \theta_{o,j}\cos(j\lambda), \quad K_n = \epsilon_n(\beta)^{-1/\beta}(\log n)^{1/\beta}$$

and $\|R_0\|^2 \leq \epsilon_n^2(\beta)(\log n)^{-2}$ so that

$$\begin{aligned} \text{tr}\left[T_n^2(b_{o,n})\right] &= \text{tr}\left[T_n^2(\tilde{\Delta}_{d_o,k_{B,n},K_n})\right] + O(\log n[K_n^{-2\beta} + K_n^{-\beta}k_{B,n}^{-\beta}]) \\ &= \text{tr}\left[T_n^2(\tilde{\Delta}_{d_o,k_{B,n},K_n})\right] + O(\epsilon_n^2(\beta)), \end{aligned} \tag{3.8}$$

where the term $O(\log n[K_n^{-2\beta} + K_n^{-\beta}k_{B,n}^{-\beta}])$ comes from the fact that

$$\left|\text{tr}\left[T_n^2(b_{o,n})\right] - \text{tr}\left[T_n^2(\tilde{\Delta}_{d_o,k_{B,n},K_n})\right]\right| \leq \text{tr}\left[T_n^2(R_0)\right] + |T_n(R_0)||T_n(\tilde{b}_{o,n})|$$

and from the use of inequality (20) in Lemma 6 of RCL, with $f_1 = f_2 = 1$, $\delta = 0$ and b either equal to $\tilde{\Delta}_{d_o,k_{B,n},K_n}$ or R_0. Note that the constant in the term $O(\epsilon_n^2(\beta))$ in (3.8) does not depend on L. Lemma 2.1 in Kruijer and Rousseau [5] together with the fact that

$$|\tilde{\Delta}_{d_o,k_{B,n},K_n}(\lambda) - \tilde{\Delta}_{d_o,k_{B,n},K_n}(y)| \leq \sum_{j=k_{B,n}}^{K_n} j|\theta_{o,j}| \leq C(\beta)L_0 K_n^{-\beta+3/2} \vee k_{B,n}^{-\beta+3/2}$$

implies that for large enough n,

$$\begin{aligned} &n^{-1}\left|\text{tr}\left[T_n^2(\tilde{\Delta}_{d_o,k_{B,n},K_n})\right] - 2\pi\text{tr}\left[T_n(\tilde{\Delta}^2_{d_o,k_{B,n},K_n})\right]\right| \\ &\quad\leq KC(\beta)L_o n^{-2\beta+1+\epsilon}\epsilon_n^2(\beta) = o(\epsilon_n^2(\beta)), \quad \forall \epsilon > 0, \end{aligned}$$

uniformly over $\beta_s \leq \beta \leq \beta_2$ and $\theta_o \in \Theta(\beta, L_o)$. Consequently,

$$\frac{c}{n}tr[T_n^2(b_{o,n})] \leq \text{tr}\left[T_n^2(\tilde{\Delta}_{d_o,k_{B,n},K_n})\right] + C(L_o,\beta)\epsilon_n^2(\beta), \tag{3.9}$$

for a constant $C(L_o, \beta)$ independent of L. Next we apply Lemma 2.4 in Kruijer and Rousseau [5] with $f = g_o$ and $b_1 = b_2 = b_{o,n}$; it then follows that

$$\gamma_1 \leq \|b_{o,n}\|_\infty^2 n^{\delta-1}n^{\epsilon-1} = o(\epsilon_n^2(\beta)), \quad \forall \epsilon > 0. \tag{3.10}$$

Finally Lemma 2.3 in Kruijer and Rousseau [5] implies that for all $\epsilon > 0$,

$$\gamma_2 \leq \|b_{o,n}\|_\infty^2 n^{-1+\epsilon} = o(\epsilon_n^2(\beta)). \tag{3.11}$$

Combining (3.9), (3.10) and (3.11), it follows that

$$h_n(f_{o,k_{B,n}}, f_o) = n^{-1}\text{tr}\left[T_n(\tilde{\Delta}^2_{d_o,k_{B,n},K_n})\right] + C(L_o,\beta)\epsilon_n^2(\beta) \leq 2C'(L_o,\beta)\epsilon_n^2(\beta),$$

where also $C'(L_o,\beta)$ is independent of L. The last inequality follows from

$$n^{-1}\mathrm{tr}\left[T_n(\tilde{\Delta}^2_{d_o,k_{B,n},K_n})\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde{\Delta}^2_{d_o,k_{B,n},K_n}(\lambda)d\lambda$$
$$= \sum_{j=k_{B,n}+1}^{K_n}\theta^2_{o,j} \le C'(L_o,\beta)\epsilon_n^2(\beta).$$

We now bound the last term in (3.6), which we write as

$$h_n(f_{o,k_{B,n}},f) = \frac{1}{2n}\mathrm{tr}\left[T_n(f_{o,k_{B,n}})^{-1}T_n(fb)T_n(f)^{-1}T_n(fb)\right], \quad b=(f-f_{o,k_{B,n}})/f.$$

Since $d\ge d_o$, $|b|_\infty < +\infty$ and applying Lemma 6 inequality (20) of RCL, we obtain if $d,\theta\in\tilde{\mathcal{B}}_n$ with $a>0$,

$$h_n(f_{o,k_{B,n}},f) \le C\log n\left(|b|_2^2+|d-d_o||b|_\infty^2\right)$$
$$\le C\log n\left(\sum_{j=1}^{k_{B,n}}(\theta_j-\theta_{o,j})^2 + n^{-a}\epsilon_n^2(\beta)\right) = o(\epsilon_n^2(\beta)),$$

which finally implies that $\pi(\bar{\mathcal{B}}_n)\ge\pi(\tilde{\mathcal{B}}_n)$ and that (3.4) is proved. We now find an upper bound on N_n. First write $\bar{W}_n = W_n\cap\mathcal{F}_n$ where $\mathcal{F}_n = \{f_{d,k,\theta}; k\le k_{B_1,n}\}$ and $k_{B_1,n} = B_1(n/\log n)^{1/(2\beta+1)}$. Then, since the prior on k is Poisson,

$$\pi(\mathcal{F}_n^c)\le e^{-c_2k_{B_1,n}\log k_{B_1,n}}\le e^{-2nu_0\epsilon_n^2(\beta)}$$

if B_1 is large enough (depending on L_o,c_2,β_s,β_2) and W_n can be replaced by $\bar{W}_n$ in the definition of N_n. Following the proof of RCL, we decompose $\bar{W}_n = \cup_{l=l_0}^{l_n}W_{n,l}$, where $l_0\ge 2$, $l_n = \lceil\epsilon^2/\epsilon_n^2(\beta)\rceil - 1$ and

$$W_{n,l} = \left\{f_{d,k,\theta}\in\mathcal{G}: k\le k_{B_1,n}, h(f_{d,k,\theta},f_o)\le\epsilon^2,\ |d-d_o|\le\delta,\right.$$
$$\left.\epsilon_n^2 l\le h_n(f_o,f_{d,k,\theta})\le u_n(l+1)\right\}.$$

In addition let $N_{n,l} = \int_{W_{n,l}}R_n(f)d\pi(f)$; then $N_n = \sum_{l=l_0}^{l_n}N_{n,l}$, and we have

$$E_o\left[\frac{N_n}{D_n}\right]\le P_o^n\left(D_n\le e^{-nu_0\epsilon_n^2(\beta)}/2\right) + E_o\left[\sum_{l=l_0}^{l_n}\frac{N_{n,l}}{D_n}1_{\{D_n\ge e^{-nu_0\epsilon_n^2(\beta)}/2\}}\right]. \tag{3.12}$$

We construct tests $\bar\phi_l$ $(l = l_0, \dots, l_n)$ and write

$$\begin{aligned} E_o &\left[\sum_{l=l_0}^{l_n} \frac{N_{n,l}}{D_n} 1_{\{D_n \geq e^{-nu_n/2}\}}(\bar\phi_l + 1 - \bar\phi_l)\right] \\ &\leq \sum_{l=l_0}^{l_n} E_o\left(\bar\phi_l\right) + 2e^{nu_n} \sum_{l=l_0}^{l_n} E_o\left[N_{n,l}(1-\bar\phi_l)\right]. \end{aligned} \tag{3.13}$$

The tests are based on a collection of spectral densities $H_{n,l} = \cup_{k=0}^{k_{B_1,n}} H_{n,l,k} \subset W_{n,l}$ defined as follows. Let D_l be a grid over $\{d : |d - d_o| \leq \delta\}$ with spacing $l\epsilon_n^2(\beta)/(\log n)$. Let $T_{l,k}$ denote the centers of hypercubes of radius $\frac{l\epsilon_n^2(\beta)}{k}$, covering $\Theta_k(\beta_s, L)$. We define $H_{n,l,k}$ as the collection of spectral densities $f_{l,i} = (2e)^{l\epsilon_n^2(\beta)} f_{d_{l,i},k,\theta^{l,i}}$, with $d_{l,i} \in D_l$ and $\theta^{l,i} \in T_{l,k}$. With every $f_{l,i}$ we associate a test

$$\phi_{l,i} = 1_{\{X'(T_n^{-1}(f_o) - T_n^{-1}(f_{l,i}))X \geq \mathrm{tr}\{I_n - T_n(f_o)T_n^{-1}(f_{l,i}) + \frac{n}{4}h_n(f_o, f_{l,i})\}\}}, \tag{3.14}$$

and set $\bar\phi_l = \max_i \phi_{l,i}$.

The set $H_{n,l}$ can be seen as a collection of upper-bracket spectral densities, since for each $f_{d,k,\theta} \in W_{n,l}$ there exists a $f_{l,i} \in H_{n,l,k}$ such that $f_{l,i} \geq f_{d,k,\theta}$, $0 \leq d_{l,i} - d \leq l\epsilon_n^2(\beta)/(\log n)$ and

$$\begin{aligned} 0 &\leq (2e)^{l\epsilon_n^2(\beta)} \exp\left\{\sum_{j=0}^{k} \theta_j^{l,i}\cos(jx)\right\} - \exp\left\{\sum_{j=0}^{k}\theta_j\cos(jx)\right\} \\ &\leq \frac{l\epsilon_n^2(\beta)}{32}(2e)^{l\epsilon_n^2(\beta)} \exp\left\{\sum_{j=0}^{k}\theta_j^{l,i}\cos(jx)\right\}. \end{aligned} \tag{3.15}$$

The cardinality of $\cup_{k=0}^{k_{B_1,n}} H_{n,l,k}$ is at most

$$\left(l^{-1}k_{B_1,n}\epsilon_n(\beta)^{-4}\right)^{k_{B_1,n}} \frac{\delta \log n}{l\epsilon_n^2(\beta)} \leq \exp\{2k_{B_1,n}\log n\} = C_{n,l}, \tag{3.16}$$

for all $l \geq 2$. To bound the right hand side of (3.13), we use (3.16) in combination with the following error bounds for each of the tests $\phi_{l,i}$. Let $f \in W_{n,l}$ and let $f_{l,i} \in H_{n,l}$ be such that (3.15) holds, $\phi_{l,i}$ being the associated test-function. Then, from equation (4.4) in RCL together with the bound (3.1) on $b_n(f_o, f_{l,i})/h_n(f_o, f_{l,i})$ which again depends on $\|\theta^{l,i}\|_1$ and L_0, we obtain that for all $0 < \alpha < 1$, there exists constants $d_1, d_2 > 0$ depending on L_o and $\|\theta^{l,i}\|_1$ such that

$$E_o \phi_{l,i} \leq e^{-d_1 n l^\alpha \epsilon_n^2(\beta)}, \quad E_f^n (1 - \phi_{l,i}) \leq e^{-d_2 n l^\alpha \epsilon_n^2(\beta)}. \tag{3.17}$$

Using (3.17) we obtain the following bound on the term $\sum_{l=l_0}^{l_n} E_o \left(\bar{\phi}_l\right)$ in (3.13):

$$\sum_{l=l_0}^{l_n} E_o \left(\bar{\phi}_l\right) \leq \sum_{l=l_0}^{l_n} C_{n,l} e^{-d_1 n l^\alpha \epsilon_n^2(\beta)} \leq e^{2k_{B_1,n} \log n} \sum_{l=l_0}^{l_n} e^{-d_1 n l^\alpha \epsilon_n^2(\beta)} \to 0,$$

as soon as $l_0 \geq \left(\frac{2B_1}{d_1 u_0}\right)^2$, choosing $\alpha = 1/2$. Using (3.17) the last term in (3.13) is

$$\sum_{l=l_0}^{l_n} E_o \left[N_{n,l}(1 - \bar{\phi}_l)\right] = \sum_{l=l_0}^{l_n} \int_{W_{n,l}} E_f^n (1 - \bar{\phi}_l) d\pi(f) \leq e^{-d_2 n l_0^{1/2} \epsilon_n^2(\beta)} \leq e^{-2n\epsilon_n^2(\beta)}$$

as soon as $d_2 l_0^{1/2} \geq 2$, i.e. $l_0 \geq 4 d_2^{-2}$. Note that the two lower bounds on l_0 depends on L_o and on $\|\theta^{l,i}\|_1$. Finally choosing, $l_0 = \max\left(4d_2^{-2}, \left(\frac{2B_1}{d_1 u_0}\right)^2, 2, u_0\right)$ we obtain

$$P^\pi \left[h_n(f, f_o) \leq l_0 \epsilon_n^2(\beta) | X^n\right] = o(n^{-1}).$$

From that we deduce a concentration rate in terms of the l norm, following RCL's argument in Appendix C. Let l_0 be an arbitrary constant and assume that $h_n(f, f_o) \leq l_0 \epsilon_n^2(\beta)$ and $f = f_{d,k,\theta}$. Then inequality (C.3) of Lemma 6 of RCL implies that

$$\frac{1}{n} \text{tr} \left[T_n(f_o^{-1}) T_n(f_o - f) T_n(f^{-1}) T_n(f_o - f)\right] \leq C_1 l_0 \epsilon_n,$$

where C_1 depends only on $\|\theta\|_1$ and on $\|\theta_o\|_1$. This implies that

$$\frac{1}{n} \text{tr} \left[T_n(f_o^{-1}(f_o - f)) T_n(f^{-1}(f_o - f))\right] \leq 2C_1 l_0 \epsilon_n^2(\beta)$$

since the difference between the two terms is of order $O(n^{-1+2a})$ $\forall a > 0$, which also implies that $h(f_o, f) \leq 3C_1 l_0 \epsilon_n^2(\beta)$, for the same reason. Since $l(f_o, f) \leq h(f_o, f)$, we finally obtain that

$$P^\pi \left[l(f, f_o) \leq 3C_1 l_0 \epsilon_n^2(\beta) | X^n\right] = o(n^{-1}).$$

To terminate the proof of Theorem 2.1, it only remains to prove that l_0 depends only on L_o, β_s. This is done using a simple re-insertion argument. Recall also that $k \le k_{B_1,n} = B_1 n \epsilon_n^2(\beta)$, where B_1 is independent of the radius L of the sobolev-ball $\Theta(\beta_s, L)$ defining the support of the prior. We start with the following observation. From Kruijer and Rousseau [4] (equation (3.5)) it follows that for fixed d and k, the minimizer of $l(f_o, f_{d,k,\theta})$ over $\mathbb{R}^{k+1}$ is

$$\bar{\theta}_{d,k} := \text{argmin}_{\theta \in \mathbb{R}^{k+1}} l(f_o, f_{d,k,\theta}) = \theta_{o[k]} + (d_o - d)\eta_{[k]},$$

where η is defined by $\eta_j = -2/j$ $(j \ge 1)$ and $\eta_0 = 0$. Assuming that $l(f_o, f_{d,k,\theta}) \le 3C_1 l_0 \epsilon_n^2(\beta)$ and $k \le k_{B_1,n}$ leads to $l(f_o, f_{d,k,\bar{\theta}_{d,k}}) \le 3C_1 l_0 \epsilon_n^2$ and $\|\theta - \bar{\theta}_{d,k}\|^2 = l(f_{d,k,\theta}, f_{d,k,\bar{\theta}_{d,k}}) \le 12 C_1 l_0 \epsilon_n^2$. Therefore

$$\begin{aligned}\sum_{j=0}^{k} |\theta_j| &\le \sum_{j=0}^{k_{B_1,n}} |\theta_j - (\bar{\theta}_{d,k_n})_j| + \sum_{j=0}^{k_{B_1,n}} |(\bar{\theta}_{d,k_n})_j| \\ &\le \sqrt{12C}\sqrt{l_0}\epsilon_n k_{B_1,n}^{1/2} + 2|d - d_o|\log n + \sum_{j=0}^{k_{B_1,n}} j^\rho |\theta_{o,j}| \le 2(2\beta_s - 1)^{-1/2}\sqrt{L_o}\end{aligned}$$

when n is large enough, where the second inequality comes from Lemma 3.1 of Kruijer and Rousseau [4]. This achieves the proof of Theorem 2.1.

Acknowledgement

This work was supported by the 800-2007–2010 grant ANR-07-BLAN-0237-01 “SP Bayes”.

References

[1] Arbel, J. (2010). Bayesian optimal adaptive estimation using a sieve prior, submitted.

[2] Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika*, 80(4):817–822.

[3] Holan, S., McElroy, T., and Chakraborty, S. (2009). A Bayesian approach to estimating the long memory parameter. *Bayesian anlysis*, 4(1):159–190.

[4] Kruijer, W. and Rousseau, J. (2011a). Bayesian semi-parametric estimation of the long-memory parameter under fexp-priors.

[5] Kruijer, W. and Rousseau, J. (2011b). Bayesian semi-parametric estimation of the long-memory parameter under fexp-priors. (supplement).

[6] Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In *Theory and applications of long-range dependence*, pages 251–301. Birkhäuser Boston, Boston, MA.

[7] Rivoirard, V. and Rousseau, J. (2010). Bernstein-von mises theorem for linear functionals of the density.

[8] Rousseau, J., Chopin, N., and Liseo, B. (2010). Bayesian nonparametric estimation of the spectral density of a long memory gaussian process.