Bayesian nonparametric estimation of the spectral
density of a long memory Gaussian time series

Judith Rousseau  Brunero Liseo*
Université Paris Dauphine  Università di Roma “La Sapienza”

Abstract
Let $X = \{X_t, t = 1, 2, \ldots \}$ be a stationary Gaussian random process, with mean $E X_t = \mu$ and covariance function $\gamma(\tau) = E(X_t - \mu)(X_{t+\tau} - \mu)$. Let $f(\lambda)$ be the corresponding spectral density; a stationary Gaussian process is said to be long-range dependent, if the spectral density $f(\lambda)$ can be written as the product of a slowly varying function $\tilde{f}(\lambda)$ and the quantity $\lambda^{-2d}$. In this paper we propose a novel Bayesian nonparametric approach to the estimation of the spectral density of $X$. We prove that, under some specific assumptions on the prior distribution, our approach assures posterior consistency both when $f(\cdot)$ and $d$ are the objects of interest. The rate of convergence of the posterior sequence depends in a significant way on the structure of the prior; we provide some general results and also consider the fractionally exponential (FEXP) family of priors (see below). Since it has not a well founded justification in the long memory set-up, we avoid using the Whittle approximation to the likelihood function and prefer to use the true Gaussian likelihood. It makes the computational burden of the method quite


Key words and Phrases: Kullback-Leibler distance, fractionally exponential priors, Population Monte Carlo, spectral analysis, Toeplitz matrices.
challenging. To mitigate the impact of that in finite sample computations, we propose to use a Population Monte Carlo (PMC) algorithm, which avoids rejecting some proposed values, as it regularly happens with MCMC algorithms. We also propose an extension of PMC in order to deal with the case of varying dimension parameter space. We finally present an application of our approach.

1 Introduction

Let $X = \{X_t, \ t=1,2,\ldots\}$ be a stationary Gaussian random process, with mean $E X_t = \mu$ and covariance function $\gamma(\tau) = E(X_t-\mu)(X_{t+\tau}-\mu)$. Let $f(\lambda)$ be the corresponding spectral density, which satisfies the relation

$$\gamma(\tau) = \int_{-\pi}^{\pi} f(\lambda)e^{it\lambda}d\lambda \quad (\tau = 0, \pm 1, \pm 2, \ldots).$$

A stationary Gaussian process is said to be long-range dependent, if there exist a positive number $C$ and a value $d (0 < d < 1/2)$ such that

$$\lim_{\lambda \to 0} C\lambda^{-2d} f(\lambda) = 1.$$

Alternatively, one can define a long memory process as one such that its spectral density $f(\lambda)$ can be written as the product of a slowly varying function $\tilde{f}(\lambda)$ and the quantity $\lambda^{-2d}$ which causes the presence of a pole of $f(\lambda)$ at the origin.

Interest in long-range dependent time series has increased enormously over the last fifteen years; Beran (1994) provides a comprehensive introduction and the book edited by Doukhan, Oppenheim and Taqqu (2003) explores in depth both theoretical aspects and various applications of long-range dependence analysis in several different disciplines, from telecommunications engineering to economics and finance, from astrophysics and geophysics to medical time series and hydrology.
Pioneering work on long memory processes is due to Mandelbrot and Van Ness (1968), Mandelbrot and Wallis (1969) and others. Fully parametric maximum likelihood estimates of $d$ were introduced in the Gaussian case by Fox and Taqqu (1986) and Dahlhaus (1989) and they have recently been developed in much greater generality by Giraitis and Taqqu (1999); a regression approach to the estimation of the spectral density of long memory time series is provided in Geweke and Porter-Hudak (1983); generalized linear regression estimates were suggested by Beran (1993). However, parametric inference can be highly biased under mis-specification of the true model: this fact has suggested semiparametric approaches: see for instance Robinson (1995).

Due to factorization of the spectral density $f(\lambda) = \lambda^{-2d} \tilde{f}(\lambda)$, a semiparametric approach to inference seems particularly appealing in this context. One needs to estimate $d$ as a measure of long-range dependence while no particular modeling assumptions on the structure of the covariance function at short ranges are necessary: Liseo, Marinucci and Petrella (2001) consider a Bayesian approach for this problem, while Bardet, Lang, Oppenheim, Philippe, Stoev and Taqqu (2003) provide an exhaustive review on the classical approaches.

Practically all the existing procedures either exploit the regression structure of the log-spectral density in a reasonably small neighborhood of the origin (Robinson 1995) or use an approximate likelihood function based on the so called Whittle’s approximation (Whittle 1962), where the original data vector $X_n = (X_1, X_2, \ldots, X_n)$ gets transformed into the periodogram $I(\lambda)$ computed at the Fourier frequencies $\lambda_j = 2\pi j/n$, $j = 1, 2, \ldots, n$, and the “new” observations $I(\lambda_1), \ldots, I(\lambda_n)$ are, under a short range dependence, approximately independent, each $I(\lambda_j)/f(\lambda_j)$ having an exponential distribution. This is for example the approach taken in Choudhuri, Ghosal and Roy (2004), which develop a Bayesian nonparametric analysis for the spectral density of a short memory time series. Unfortunately, the Whittle’s approximation fails to hold in the presence of long range dependence, at least for
the smallest Fourier frequencies.

In this paper we propose a Bayesian nonparametric approach to the estimation of the spectral density of the stationary Gaussian process: we avoid the use of the Whittle approximation and we deal with the true Gaussian likelihood function.

The literature on Bayesian nonparametric inference has increased tremendously in the last decades, both from a theoretical and a practical point of view. Much of this literature has dealt with the independent case, mostly when the observations are identically distributed. The theoretical perspective was mainly dedicated to either construction of processes used to define the prior distribution with finite distance properties of the posterior, in particular when such a prior is conjugate, see for instance Ghosh and Ramamoorthi (2003) for a review on this, or to consistency and rates of convergence properties of the posterior, see for instance Ghosal, Ghosh and van der Vaart (2000) or Shen and Wasserman (2001).

The dependent case has hardly been considered from a theoretical perspective apart from Choudhuri et al. (2004), who deal with Gaussian weakly dependent data and, in a more general setting, Ghosal and Van der Vaart (2006). In this paper we study the asymptotic properties of the posterior distributions for Gaussian long-memory processes, where the unknown parameters are the spectral density and the long-memory parameter $d$. General consistency results are given and a special type of prior, namely the FEXP prior as it is based on the FEXP model, is studied. From this, consistency of Bayesian estimators of both the spectral density and the long memory parameter are obtained. To understand better the link between the Bayesian and the frequentist approaches we also study the rates of convergence of the posterior distributions, first in a general setup and then in the special case of FEXP priors. The approach considered here is similar to what is often used in the independent and identically distributed case, see for instance Ghosal et al. (2000). In particular we need to control prior probability on some neighborhood of the true spectral
density and to control a sort of entropy of the prior (see Section 3); however the techniques are quite different due to the dependence structure of the process.

The gist of the paper is to provide a fully nonparametric Bayesian analysis of long range dependence models. In this context there already exist many elegant and maybe more general (in the sense of being valid even without the Gaussian assumption) classical solutions. However we believe that a Bayesian solution would be still important because of the following reasons.

i) By definition, our scheme allows to include in the analysis some prior information which may be available in some applications.

ii) While classical solutions are, in a way or another, based on some asymptotic arguments, our Bayesian approach relies only on the observed likelihood function (and prior information).

iii) We are able to provide a valid approximation to the “true” posterior distribution of the main parameters of interest in the model, namely the long memory parameter \( d \) or the global spectral density.

Also, on a more theoretical perspective, we believe that this paper can be useful to clarify the intertwines between Bayesian and frequentist approaches to the problem. We also present a specific algorithm to implement the procedure, i.e. to simulate from the posterior or approximately so. The algorithm used is a version of the Population Monte-Carlo algorithm as devised in Douc, Guillin, Marin and Robert (2006). Although this is not the main focus of the paper, the computation of the posterior distribution is an important issue since the likelihood is difficult to calculate: all the details about the practical implementation of the algorithm can be found in Liseo and Rousseau (2006).
The paper is organized as follows: in the next section we first introduce the necessary notation and mathematical objects; then we provide a general theorem which states some sufficient condition to ensure consistency of the posterior distribution. We also discuss in detail a specific class of priors, the FEXP prior, which takes its name after the fractional exponential model which has been introduced by Robinson 1991, 1994 to model the spectral density of a covariance stationary long-range dependent process. The FEXP model can be seen as a generalization of the exponential model proposed by Bloomfield (1973) and it allows for semi-parametric modeling of long range dependence; see also Beran (1994) or Hurvich, Moulines and Soulier (2002). In Section 3 we study the rate of convergence of the posterior distribution first in the general case and then in the case of FEXP priors and in Section 4 we give details about computational issues. The final section is devoted to some discussion.

### 2 Consistency results

We observe a set of \( n \) consecutive realizations \( X_n = (X_1, \ldots, X_n) \) from a Gaussian stationary process with spectral density \( f_0 \), where \( f_0(\lambda) = |\lambda|^{-2d_0} \tilde{f}_0(\lambda) \). Because of the Gaussian assumption, the density of \( X_n \) can be written as

\[
\varphi_{f_0}(X_n) = \frac{e^{-\frac{1}{2}X'_n T_n(f_0)^{-1}X_n}}{|T_n(f_0)|^{1/2}(2\pi)^{n/2}},
\]

where \( T_n(f_0) = [\gamma(j - k)]_{1 \leq j, k \leq n} \) is the covariance matrix with a Toeplitz structure. The aim is to estimate both \( \tilde{f}_0 \) and \( d_0 \) using Bayesian nonparametric methods.

Let \( \mathcal{F} = \{ f, f \text{ symmetric on } [-\pi, \pi], \int |f| < \infty \} \) and \( \mathcal{F}_+ = \{ f \in \mathcal{F}, f \geq 0 \} \); then \( \mathcal{F}_+ \) denotes the set of spectral densities. We first define three types of pseudo-distances on \( \mathcal{F}_+ \).
The Kullback-Leibler divergence for finite $n$ is defined as

$$KL_n(f_0; f) = \frac{1}{n} \int_{\mathbb{R}^n} \varphi_{f_0}(X_n) \left[ \log \varphi_{f_0}(X_n) - \log \varphi_f(X_n) \right] dX_n$$

$$= \frac{1}{2n} \left[ \text{tr} \left( T_n(f_0) T_n^{-1}(f) - \text{id} \right) - \log \det(T_n(f_0) T_n^{-1}(f)) \right]$$

where $\text{id}$ represents the identity matrix of the appropriate order. Letting $n \to \infty$, we can define, when it exists, the quantity

$$KL_\infty(f_0; f) = \frac{1}{\pi} \int_{-\pi}^{\pi} \left[ f_0(\lambda) f(\lambda) - 1 - \log \frac{f_0(\lambda)}{f(\lambda)} \right] d\lambda.$$  

We also define two symmetrized version of $KL_n$, namely

$$h_n(f_0, f) = KL_n(f_0; f) + KL_n(f; f_0); \quad d_n(f_0, f) = \min\{KL_n(f_0; f), KL_n(f; f_0)\}$$

and their corresponding limits as $n \to \infty$:

$$h(f_0, f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \frac{f_0(\lambda)}{f(\lambda)} + \frac{f(\lambda)}{f_0(\lambda)} - 2 \right] d\lambda; \quad d(f_0, f) = \min\{KL_\infty(f_0; f), KL_\infty(f; f_0)\}.$$  

We also consider the $L^2$ distance between the logarithms of the spectral densities, namely

$$\ell(f, f') = \int_{-\pi}^{\pi} (\log f(\lambda) - \log f'(\lambda))^2 d\lambda. \quad (2)$$

This distance has been considered in particular by Moulines and Soulier (2003). This is quite a natural distance in the sense that it always exists, whereas the $L^2$ distance between $f$ and $f'$ need not, at least in the types of models considered in this paper. Let $\pi$ be a prior probability distribution on the set

$$\tilde{\mathcal{F}} = \{ f \in \mathcal{F}, f(\lambda) = |\lambda|^{-2d}\tilde{f}(\lambda), \quad \tilde{f} \in C^0, -\frac{1}{2} < d < \frac{1}{2} \}, \quad \tilde{\mathcal{F}}_+ = \{ f \in \tilde{\mathcal{F}}_+, f \geq 0 \},$$

where $C^0$ is the set of continuous functions on $[-\pi, \pi]$. Let $A_\varepsilon = \{ f \in \tilde{\mathcal{F}}_+; d(f, f_0) \leq \varepsilon \}$. Our first goal will be to prove the consistency of the posterior distribution of $f_0$, that is, we want to show that

$$P^\pi[A_\varepsilon|X_n] \to 0, \quad f_0 \text{ a.s.}$$
From this, we will be able to deduce the consistency of some Bayes estimators of the spectral density \( f \) and of the long memory parameter \( d \). We first state and prove the strong consistency of the posterior distribution under very general conditions both on the prior and on the true spectral density. Then, building on these results, we will obtain the consistency of a class of Bayes estimates of the spectral density, together with the consistency of the Bayes estimates of the long memory parameter \( d \). The already introduced FEXP class of prior will be then proposed, and its use will be explored in detail.

### 2.1 The main result

In this section we derive the main result about consistency of the posterior distribution. We also discuss the asymptotic behavior of the posterior point estimates of some parameter of major interest, such as the long memory parameter \( d \) and the global spectral density.

Consider the following two subsets of \( \mathcal{F} \)

\[
\mathcal{G}(d, M, m, L, \rho) = \left\{ f \in \tilde{\mathcal{F}}_+: f(\lambda) = |\lambda|^{-2d} \tilde{f}(\lambda), m \leq \tilde{f}(\lambda) \leq M, \left| \tilde{f}(x) - \tilde{f}(y) \right| \leq L|x - y|^\rho \right\},
\]

where \(-1/2 < d < 1/2, m, M, \rho > 0\);

\[
\mathcal{F}(d, M, L, \rho) = \left\{ f \in \tilde{\mathcal{F}}; f(\lambda) = |\lambda|^{-2d} \tilde{f}(\lambda), |\tilde{f}(\lambda)| \leq M, |\tilde{f}(x) - \tilde{f}(y)| \leq L|x - y|^\rho \right\}.
\]

The boundedness constraint on \( \tilde{f} \) in the definition of \( \mathcal{G}(d, M, m, L, \rho) \) is necessary here to guarantee the identifiability of \( d \), while the Lipschitz-type condition on \( \tilde{f} \), in both definitions, are actually needed to ensure that normalized traces of products of Toeplitz matrices, that typically appear in the distances considered previously, will converge. We also consider the following set of spectral densities, which is of interest in the study of rates of convergence:
let

\[ L^\star(M, m, L) = \{ h(\cdot) \geq 0, 0 < m \leq h(\cdot) \leq M, |h(x) - h(y)| \leq L|x - y|(|x| \wedge |y|)^{-1} \} \]

and

\[ L(d, M, m, L) = \{ f = |\lambda|^{-2d} \hat{f}(\lambda), \hat{f} \in L^\star(M, m, L) \}. \]

Note that \( G \) and \( L \) are similar, with only a slight modification on the Lipschitz condition. The set \( L \) has been considered in particular in Moulines and Soulier (2003).

We now consider the main result on the consistency of the posterior distribution

**Theorem 1** Let \( \bar{G}(t, M, m, L, \rho) = \bigcup_{0 \leq d \leq 1/2-t} \mathcal{G}(d, M, m, L, \rho) \), and assume that there exists \((t_0, M_0, m_0, L_0)\) such that we have either \( f_0 \in \bar{G}(t, M_0, m_0, L_0, \rho_0) \) with \( 1 \geq \rho_0 > 0 \) or \( f_0 \in \bigcup_{0 \leq d \leq 1/2-t_0} L(d, M_0, m_0, L_0) \). Let

\[ \bar{F}_+(t, M, m) = \bigcup_{-1/2+t \leq d \leq 1/2-t} \{ f \in F_+, f(\lambda) = |\lambda|^{-2d} \hat{f}(\lambda), 0 < m \leq \hat{f} \leq M \}. \]

Let \( \pi \) be a prior distribution such that

i) \( \forall \varepsilon > 0 \) and for some \( M' > 0 \), there exists \( M, m, L, \rho > 0 \) such that if

\[ B_\varepsilon = \left\{ f \in \bar{G}(t, M, m, L, \rho) : h(f_0, f) \leq \varepsilon, 6(d_0 - d) < \rho_0 \wedge \frac{1}{2}, \int \left( \frac{f_0}{f} - 1 \right)^3 dx \leq M' \right\}, \]

then \( \pi(B_\varepsilon) > 0 \). For simplicity, in our notations the case

\[ f_0 \in \bigcup_{0 \leq d \leq 1/2-t_0} L(d, M_0, m_0, L_0) \]

corresponds to \( \rho_0 = 1 \). This simplification is also used in (ii).

ii) \( \forall \varepsilon > 0 \), small enough, there exists \( F_n \subset \{ f \in F_+, d(f_0, f_i) > \varepsilon \} \), such that \( \pi(F_n) \leq e^{-nr} \) and there exist \( t, M, m, C > 0 \) with \( t < \rho_0/4 \), and a smallest possible net \( \mathcal{H}_n \subset \{ f \in \bar{F}_+(t, M, m); d(f, d_0) > \varepsilon/2 \} \) such that when \( n \) is large enough, \( \forall f \in F_n, \exists f_i \in \)
Denote by $H_n$ the logarithm of the cardinality of the smallest possible net $H_n$. Then, if

$$N_n \leq nc_1, \quad \text{with} \quad c_1 < \varepsilon \log \varepsilon^{-1}/2, \quad 0 < \delta$$

then

$$P^\pi[A_\varepsilon|X_n] \to 1, \quad f_0 \text{ a.s.} \quad (5)$$

Proof. See Appendix B. ■

The above theorem is important to clarify which conditions on the prior distribution $\pi$ are really crucial in a long memory setting, where the techniques usually adopted in the i.i.d. case, cannot be used and even the adoption of a Whittle approximation is not legitimate in this setting (at least at the lowest frequencies). From a practical perspective, however, the hardest part of the job is actually to verify whether a specific type of priors actually meets the conditions listed in Theorem 1. Just to mention two difficulties, the construction of the net $H_n$ in the proof of Theorem 1 may be strongly dependent on the prior we use; also it may depend, in a non trivial way, on the sample size. To be more precise, checking the uniform bound on the terms in the form $n^{-1} \text{tr} (\{T_n(|\lambda|^{-2d_0})^{-1}T_n(f_i - f_0)\}^2)$ might be quite delicate. In Appendix B, we also give a more general set of conditions to obtain consistency which is however quite cumbersome but might prove to be useful in some situations.

We will discuss in detail these issues in the context of the FEXP prior in §2.3.
2.2 Consistency of estimates for some quantities of interest

We now discuss the problem of consistency for the Bayes estimates of the spectral density. The quadratic loss function on $f$ is not a natural loss function for this problem, since there exist some spectral densities in $F$ that are not square integrable (if $d > 1/4$). A more reasonable loss function may be the quadratic loss on the logarithm of $f$, as defined by (2), which is always integrable. The Bayes estimator of $f$ associated with the loss $\ell$ and the prior $\pi$ is given by

$$\hat{f}(\lambda) = \exp\{E^\pi[\log f(\lambda)|X_n]\}.$$ 

Also, in many applications, the real parameter of interest is just $d$, the long memory exponent. It is possible to deduce, from Theorem 1, that the posterior mean of $d$, that is the Bayes estimator associated with the quadratic loss on $d$, is actually consistent.

**Corollary 1** Under the assumptions of Theorem 1, for all $\epsilon > 0$, as $n \to \infty$,

$$\pi\left\{ f = |\lambda|^{-2d}\tilde{f}; |d - d_0| > \epsilon \right\} |X_n| \to 0 \quad f_0 \ a.s$$

and $\hat{d} = E^\pi[d|X_n] \to d_0, \ f_0 \ a.s.$

**Proof.** The result comes from the fact that, when $|d - d_0| > \epsilon$, both $KL_\infty(f; f_0)$ and $KL_\infty(f_0; f)$ are greater than some fixed value $\epsilon'$ depending on $\epsilon$ only. Indeed, let $f(\lambda) = |\lambda|^{-2d}\tilde{f}(\lambda)$ and $f_0(\lambda) = |\lambda|^{-2d_0}\tilde{f}_0(\lambda)$, for all $0 < \tau < \pi$.

$$KL_\infty(f; f_0) > 2 \int_0^\tau \left[ \frac{\tilde{f}(\lambda)}{\tilde{f}_0(\lambda)} \lambda^{-2(d-d_0)} - 1 + 2(d - d_0) \log \lambda - \log (\tilde{f}/\tilde{f}_0)(\lambda) \right] d\lambda$$

$$\geq 2 \left( \frac{m}{M(1 - 2(d - d_0))} \right)^{1 - 2(d - d_0)} - \tau(1 + \log M/m + 2(d - d_0))$$

$$+ 2(d - d_0)\tau \log \tau \geq \epsilon'$$

11
where $\epsilon'$ is based on either $\tau^{1-2(d-d_0)}$ if $d > d_0$ or on $(d_0 - d)\tau \log 1/\tau$ if $d < d_0$, when $\tau$ is small enough (but fixed, depending on $\epsilon, m, M$). This implies that

$$\pi[A_{\epsilon'}^c|X] \geq \pi \{f = |\lambda|^{-2\tilde{d}f}; |d - d_0| > \epsilon\}|X| \to 0, \quad f_0 \text{ a.s.}$$

Since $d$ is bounded, a simple application of the Jensen’s inequality gives

$$(d - d_0)^2 \leq E\pi[(d - d_0)^2|X] \to 0, \quad f_0 \text{ a.s.}$$

It is also possible to derive consistency results for the point estimate of the whole spectral density:

**Corollary 2** Under the assumptions of Theorem 1, as $n \to \infty$,

$$\ell(f_0, \hat{f}) \to 0, \quad f_0 \text{ a.s.}$$

**Proof.** The idea is to prove that for all $\epsilon > 0$, there exists $\epsilon' > 0$ such that $l(f, f_0) > \epsilon$ implies that $d(f, f_0) > \epsilon'$. Indeed, when $c$ is small enough, there exists $x_c < 0$ ($x_c$ goes to $-\infty$ when $c$ goes to 0) such that $e^x - 1 - x \leq cx^2$. Then

$$KL_{\infty}(f; f_0) \geq c \left( l(f, f_0) - \int_{f/f_0 < e^{x_c}} (\log f(x) - \log f_0(x))^2 dx \right).$$

Moreover,

$$\frac{f(x)}{f_0(x)} = |x|^{-2(d-d_0)} \frac{\tilde{f}(x)}{\tilde{f}(x)} \geq \frac{m}{M} \pi^{-2(d-d_0)},$$

when $d > d_0$. Hence, by choosing $c$ small enough, the set $\{f/f_0 < e^{x_c}\} \cap \{d > d_0\}$ is empty. In this case the set $\{f/f_0 < e^{x_c}\}$ is a subset of $\{|x| \leq a_c\}$, where $a_c$ goes to zero as $c$ goes to zero; when $c \to 0$,

$$\int_{|x| \leq a_c} (\log f(x) - \log f_0(x))^2 dx \leq 2a_c \left( 4(d - d_0)^2(\log (a_c)^2 - 2\log (a_c)) + (\log M/m)^2 \right) \to 0,$$
therefore by choosing $c$ small enough there exists $\epsilon'$ such that $d(f, f_0) > \epsilon'$. This implies that $\pi \left[ \tilde{A}_c^* | X_n \right] \to 0$, when $n$ goes to infinity, $f_0$ almost surely. Using Jensen’s inequality this implies in particular that $l(f_0, \hat{f}) \to 0$, $f_0$ a.s. ■

Since the conditions stated in Theorem 1 are somewhat non standard, they need to be carefully checked for the specific class of priors one is dealing with. Here we consider the class of Fractionally Exponential priors (FEXP), and we show that these priors actually fulfill the above conditions.

2.3 The FEXP prior

Consider the set of the spectral densities with the form

$$f(\lambda) = |1 - e^{i\lambda}|^{-2d} \tilde{f}(\lambda),$$

where $\log \tilde{f}(x) = \sum_{j=0}^{K} \theta_j \cos(jx)$, and assume that the true log spectral density satisfies $\log \tilde{f}_0(x) = \sum_{j=0}^{\infty} \theta_{0j} \cos(jx)$ (in other words, it is equal to its Fourier series expansion), with

$$|\tilde{f}_0(x) - \tilde{f}_0(y)| \leq L \frac{|x - y|}{|x| \wedge |y|}, \sum_j |\theta_{0j}| < \infty,$$

for all $x$ and $y$ in $[-\pi, \pi]$. This construction has been considered, from a frequentist perspective, in Hurvich et al. (2002). Note that there exists an alternative and equivalent way of writing a FEXP spectral density in which the first coefficient of the series expansion $\theta_0$ is explicitly expressed in terms of the variance of the process, that is $\sigma^2 = 2\pi e^{\theta_0}$. We will use both the parameterizations according to notational convenience. A prior distribution on $f$ can then be expressed as a prior on the parameters $(d, K, \theta_0, ..., \theta_K)$ in the form $p(K) \pi(d|K) \pi(\theta|d, K)$, where $\theta = (\theta_0, ..., \theta_K)$, and $K$ represents the (random) order of the FEXP model. Usually, $d$ is set independent of $\theta$ for given $K$ and it is also independent of $K$ itself. Let $\pi(d) > 0$ on $[-1/2 + t, 1/2 - t]$, for some $t > 0$, arbitrarily small. Let $K$
be a priori Poisson distributed and, conditionally on $K$, notice that $\pi(\theta|K)$ needs to put mass 1 on the set of $\theta$’s such that $\sum_{j=0}^{K} |\theta_j| \leq A$, for some value $A$ large but finite. A possible way to formalize it, is to assume that, for given $K$, the quantity $S_K = \sum_{j} |\theta_j|$ has a finite support distribution; then, setting $V_j = |\theta_j|/S_K$, $j = 1, \ldots, K$, one may consider a distribution on the set $\{z \in \mathbb{R}^K; z = (z_1, \ldots, z_K), \sum z_i = 1, z_i \geq 0\}$ for example:

$$(V_1, \ldots, V_K) \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_K).$$

Since the variance of the $|\theta_j|$’s should be decreasing, we may assume, for example, that, for all $j$’s, $\alpha_j = O((1 + j)^{-2})$. Note that if we further assume that $S_K$ has a Gamma distribution with mean $\sum_{j} \alpha_j$ and variance $\sum_{j} \alpha_j^2$ then we are approximately assuming (modulo the truncation at $A$) that $|\theta_1|, \ldots, |\theta_K|$ are independent Gamma$(1, \alpha_j)$ random variables. Alternative parameterization are also available here; for example one can assume that $(V_1, \ldots, V_K)$ follows a logistic normal distribution (Aitchison and Shen 1980), which allows for a more flexible elicitation. Under the above conditions on the prior, the posterior distribution is strongly consistent, in terms of the distance $d(\cdot, \cdot)$, the estimator $\hat{f}$ as described in the previous section is almost surely consistent and so is the estimator $\hat{d}$. To prove this, we prove that the FEXP prior satisfies assumptions (i) and (ii). First, we check assumption (i): let $K_\epsilon$ be such that $\sum_{j=K_\epsilon+1}^{\infty} |\theta_{0j}| \leq \sqrt{\epsilon/2}$, then $KL_\infty(f_0; f_{0\epsilon}) \leq \epsilon/2$, where

$$f_{0\epsilon} = (1 - e^{i\lambda})^{-2d_0} \exp \left\{ \sum_{j=0}^{K_\epsilon} \theta_{0j} \cos j\lambda \right\}.$$ 

Let $\theta = (\theta_0, ..., \theta_{K_\epsilon})$ be such that $|\theta_{0j} - \theta_j| \leq \sqrt{\epsilon/4a|\theta_{0j}|}$, $j = 1, \ldots, K_\epsilon$, where $a = \sum_{j} |\theta_{0j}|$.

If $|d - d_0| < \tau$, with $\tau$ small enough, then $k_\infty(f_0, f) \leq \epsilon$. Obviously $\pi_K(\{\theta : |\theta_j - \theta_{0j}| < \sqrt{\epsilon/4a|\theta_{0j}|}, \forall j \leq K\}) > 0$, as soon as $A > \sum_{j} |\theta_{0j}|$. Moreover, for each $f(\lambda) = (1 - e^{i\lambda})^{-2d} \exp\{\sum_{j=0}^{K} \theta_j \cos (j\lambda)\}$, provided that the parameters

14
satisfy the above constraint, one has

\[
\frac{f_0(\lambda)}{f(\lambda)} = |1 - e^{i\lambda}|^{-2(d_0 - d)} \exp \left\{ \sum_j (\theta_{0j} - \theta_j) \cos (j\lambda) \right\} \leq 2|1 - e^{i\lambda}|^{-2(d_0 - d)}
\]

so that

\[
\int (f_0/f - 1)^2 d\lambda \leq M'
\]

for some constant \( M' > 0 \). Now we verify assumption (ii). Let \( \epsilon > 0 \) and set \( f_{k,d,\theta}(\lambda) = |1 - e^{-i\lambda}|^{-2d} \exp \{ \sum_{j=0}^k \theta_j \cos (j\lambda) \} \); consider

\[
\mathcal{F}_n = \{ f_{k,d,\theta}, \ d \in [-1/2 + t, 1/2 - t], k \leq k_n \},
\]

where \( k_n = k_0 n / \log n \). Since \( \pi(K \geq k_n) < e^{-nr} \), for some \( r \) depending on \( k_0 \), we have that \( \pi(\mathcal{F}_n \geq k_n) < e^{-nr} \). Now consider spectral densities in the form,

\[
f_i(\lambda) = (1 - \cos \lambda)^{-d - \delta_1} \exp \left\{ \sum_{j=0}^k \theta_j \cos j\lambda + \delta_2 \right\},
\]

where \( 0 < \delta_1, \delta_2 < c\epsilon |\log \epsilon|^{-1} \) for some constant \( c > 0 \). We prove that if \( |d' - d| < \delta_1/2 \) and \( |\theta'_j - \theta_j| < c' \delta_2 [(j + 1) \log (j + 1)^2]^{-1} \), where \( c' = \sum_{j \geq 1} j^{-1} \log j^2 \), then \( f' = f_{k,d',\theta'} \leq f_i \) and

\[
0 \leq f_i(\lambda) - f'(\lambda) \leq C f_i(\lambda) |\log (1 - \cos \lambda)| [\delta_1 + \delta_2] \leq \epsilon |\log \epsilon|^{-1} |\lambda|^{-2d_i - t/2},
\]

for some constant \( C > 0 \). To achieve the proof of assumption (ii) of Theorem 1 we need to verify

\[
\begin{align*}
n^{-1} \text{tr} \left( T_n(|\lambda|^{-2d_i})^{-1} T_n(f_i - f_0) \right)^2 & \geq B \int (f_0/f_i - 1)^2(x) dx \\
n^{-1} \text{tr} \left( T_n(|\lambda|^{-2d_0})^{-1} T_n(f_i - f_0) \right)^2 & \geq B \int (f_i/f_0 - 1)^2(x) dx,
\end{align*}
\]

(6)

this is proved in Appendix C. The number of such upper bounds is bounded by

\[
CK_n \epsilon^{-1} \left( \frac{\epsilon}{15K_n} \right)^{-K_n},
\]
so that

\[ N_n \leq k_0 Cn \left[ \frac{1}{\log n} + 1 \right] \leq nc_1, \]

by choosing \( k_0 \) small enough and \( n \) large enough: this proves that the posterior distribution associated with the FEXP prior is actually consistent.

3 Rates of convergence

In this Section we first give a general Theorem relating rates for the posterior distribution to conditions on the prior. These conditions are, in essence, similar to the conditions obtained in the i.i.d. case, in other words there is a condition on the the prior mass of Kullback-Leibler neighborhoods of the true spectral density and an entropy condition on the support of the prior. We then present the results in the case of the FEXP prior.

3.1 Main result

We now present the general Theorem on convergence rates for the posterior distribution.

**Theorem 2** Let \((\rho_n)_n\) be a sequence of positive numbers decreasing to zero, and \(B_n(\delta)\) a ball belonging to \(\bar{G}(t, M, L, \rho) \cup \bar{L}(t, M, m, L)\), defined as

\[ B_n(\delta) = \{ f(x) = |x|^{-2(d-d_0)} \tilde{f}(x); KL_n(f_0; f) \leq \rho_n/2, |d - d_0| \leq \delta \}, \]

for some \( \rho \in (0, 1] \). Let \( \pi \) be a prior such that conditions (i) and (ii) of Theorem 1 are satisfied and assume that:

(i). There exists \( \delta > 0 \) such that \( \pi(B_n(\delta)) \geq \exp\{-n\rho_n/2\} \).

(ii). For all \( \epsilon > 0 \) small enough, there exists a positive sequence \((\epsilon_n)_n\) decreasing to zero and \( \bar{F}_n \subset \bar{F}_+ \cap \{ f, d(f, f_0) \leq \epsilon \} \), such that \( \pi(\bar{F}_n \cap \{ f, d(f, f_0) \leq \epsilon \}) \leq e^{-2n\rho_n} \).
(iii). Let

\[ S_{n,j} = \{ f \in \mathcal{F}_n : \varepsilon_n^2 j \leq h_n(f_0, f) \leq \varepsilon_n^2 (j + 1) \}, \]

with \( J_n \geq j \geq J_0, \) with \( J_0 > 0 \) fixed and \( J_n = [\varepsilon_n^2 / \varepsilon_n^2] \) \( \forall J_0 \leq j \leq J_n, \) there exists a smallest possible net \( \mathcal{H}_{n,j} \subset S_{n,j} \) such that \( \forall f \in S_{n,j}, \) \( \exists f_i \geq f \in \mathcal{H}_{n,j} \) satisfying

\[ \text{tr} \left( T_n(f_i - \text{id}) / n \right) \leq h_n(f_0, f_i) / 8, \quad \text{tr} \left( T_n(f_i - f) T_n^{-1}(f_0) \right) / n \leq h_n(f_0, f_i) / 8. \]

Denote by \( \tilde{N}_{n,j} \) the logarithm of the cardinality of the smallest possible net \( \mathcal{H}_n. \)

\[ \tilde{N}_{n,j} \leq n \varepsilon_n^2 j^\alpha, \quad \text{with} \quad \alpha < 1. \]

Then, there exist \( M, C, C' > 0 \) such that if \( \rho_n \leq \varepsilon_n^2 \)

\[ E_n^0 \left[ \pi \left( f : h_n(f_0, f) \geq M \varepsilon_n^2 | X \right) \right] \leq \max \left( e^{-n \varepsilon_n^2 C}, \frac{C'}{n^2} \right). \quad (7) \]

**Proof.** Throughout the proof \( C \) denotes a generic constant. We have

\[
\pi \left( f : h_n(f_0, f) \geq M \varepsilon_n^2 | X \right) = \frac{\int_{f : h_n(f_0, f) \geq M \varepsilon_n^2} \varphi_f(X_n) / \varphi_{f_0}(X_n) d\pi(f)}{\int \varphi_f(X_n) / \varphi_{f_0}(X_n) d\pi(f)} \leq \frac{\int_{f : h_n(f_0, f) \geq \varepsilon} \varphi_f(X_n) / \varphi_{f_0}(X_n) d\pi(f)}{\int \varphi_f(X_n) / \varphi_{f_0}(X_n) d\pi(f)} \leq \frac{\mathcal{N}_n}{\mathcal{D}_n} + R_{n,2},
\]

for some \( \varepsilon > 0. \) Theorem 1 implies that \( P_0 \left[ R_{n,2} > e^{-n \delta} \right] \leq \frac{C}{n^2}, \) for some constants \( C, \delta > 0. \)

We now consider the first term of the right hand side of the above equation. Working similarly to before, if

\[ \tilde{N}_{n,j} = \int_{f : \varepsilon_n^2 j \leq h_n(f_0, f) \leq \varepsilon_n^2 (j + 1)} \frac{\varphi_f(X_n)}{\varphi_{f_0}(X_n)} d\pi(f) \]

\[ E_n^0 \left[ \frac{\tilde{N}_{n,j}}{\mathcal{D}_n} \right] \leq \sum_{j \geq M} E_n^0 \left[ \varphi_{n,j} \right] + E_n^0 \left[ (1 - \varphi_{n,j}) \frac{\tilde{N}_{n,j}}{\mathcal{D}_n} \right], \]


where $\varphi_{n,j} = \max_{f_i \in \tilde{H}_{n,j}} \varphi_i$, and $\varphi_i$ is defined as in the previous Section:

$$\varphi_i = \frac{1}{\| X_n'(T_n^{-1}(f_i) - T_n^{-1}(f_0))X_n \|} \geq \frac{1}{\| \text{id} - T_n(f_0)T_n^{-1}(f_i) \| + h_n(f_0,f_i)/4}.$$ 

Then, (22) implies that

$$E_n^0[\varphi_{n,j}] \leq \sum_{i : f_i \in \tilde{H}_{n,j}} e^{-Cn\varepsilon^2_j} \leq \tilde{N}_{n,j} e^{-Cn\varepsilon^2_j} \leq e^{-Cn\varepsilon^2_j}.$$

We also have that

$$E_n^0[(1 - \varphi_{n,j}) \frac{N_{n,j}}{D_n}] \leq P_n^0[D_n \leq e^{-n\rho_n}] + e^{n\rho_n} \pi(F_n^c \cap \{ f : d(f, f_0) \leq \varepsilon \}) + e^{n\rho_n} \int_{S_{n,j}} E_n[f] (1 - \varphi_{n,j}) d\pi(f) \leq e^{-n\rho_n} + e^{n\rho_n} e^{-nC\varepsilon^2_j} + P_n^0[D_n \leq e^{-n\rho_n}].$$

Moreover, using the same calculations as in the proof of theorem 1

$$P_n^0[D_n \leq e^{-n\rho_n}] \leq \frac{1}{n^2},$$

and Theorem 2 is proved.

The conditions given in Theorem 2 are similar in spirit to those considered for rates of convergence of the posterior distribution in the i.i.d. case. The first one is a condition on the prior mass of Kullback-Leibler neighborhoods of the true spectral density, the second one is needed to allow for sets with infinite entropy (some kind of non compactness) and the third one is an entropy condition. The inequality (7) obtained in Theorem 2 is non asymptotic, in the sense that it is valid for all $n$. However, the distances considered in Theorem 2 heavily depend on $n$ and, although they express the impact of the differences between $f$ and $f_0$ on the observations, they are not of great practical use. The entropy condition is therefore
awkward and cannot be directly transformed into some more common entropy conditions. To state a result involving distances between spectral densities that would be more useful, we consider the special case of FEXP priors, as defined in Section 2.3. We can then obtain rates of convergence in terms of the $L_2$ distance between the log of the spectral densities, $l(f, f')$. The rates obtained are the optimal rates up to a $\log n$ term, at least on certain classes of spectral densities. It is to be noted that the calculations used when working on these classes of priors are actually more involved than those used to prove Theorem 2. This is quite usual when dealing with rates of convergence of posterior distributions, however this is emphasized here by the fact that distances involved in Theorem 4 are strongly dependent on $n$. The method used in the case of the FEXP prior can be extended to other types of priors.

3.2 The FEXP prior - rates of convergence

In this Section we apply Theorem 2 to the FEXP priors. Recall that they are defined through a parameterization based on the FEXP models. In other words, $f(\lambda) = |1 - e^{i\lambda}|^{-2d}\tilde{f}(\lambda)$, and $\log \tilde{f}(\lambda) = \sum_{j=0}^{K} \theta_j \cos j\lambda$. Then the prior can be written in terms of a prior on $(d, K, \theta_0, ..., \theta_K)$.

Define now the classes of spectral densities

$$S(\beta, L_0) = \{h \geq 0; \log h \in L^2[-\pi, \pi], \log h(x) = \sum_{j=0}^{\infty} \theta_j \cos jx, \sum_j \theta_j^2 (1 + j)^{2\beta} \leq L_0\},$$

with $\beta > 0$, and assume that there exists $\beta > 0$ such that $\tilde{f}_0 \in \mathcal{L}^*(M, m, L) \cap S(\beta, L_0)$. We can then write $f_0$ as

$$f_0(\lambda) = |1 - e^{i\lambda}|^{-2d_0} \exp \left\{ \sum_{j=0}^{\infty} \theta_j \cos j\lambda \right\}.$$
Note that $\beta$ is a smoothness parameter. These classes are considered by Moulines and Soulier (2003). We now describe the construction of the FEXP prior, so that it can be adapted to $S(\beta, L_0)$. Let $S_K$ be a r.v. with density $\sim g_A$, positive in the interval $[0, A]$, let $\eta_j = \theta_j j^\beta$ and suppose that the prior on $(\eta_1/S_K, ..., \eta_K/S_K)$ has positive density on the set $\tilde{S}_{K+1} = \{x = (x_1, ..., x_{K+1}) ; \sum_{j=1}^{K+1} x_j^2 = 1\}$. We denote this class as the class of FEXP($\beta$) priors.

We now give the rates of convergence associated with the FEXP($\beta$) priors, when the true spectral density belongs to $S(\beta, L_0)$.

**Theorem 3** Assume that there exists $\beta > 1/2$ such that $\tilde{f}_0 \in \mathcal{L}(e^{L_0}, e^{-L_0}, L) \cap S(\beta, L_0)$. Let $\pi$ be a FEXP($\beta$) prior and assume that i) $K$ follows a Poisson distribution, ii) the prior on $d$ is positive on $[-1/2 + t, 1/2 - t]$, with $t > 0$, iii) the prior $g_A$ on $S_K$ is such that $A^2 \geq L_0$. Then there exist $C, C' > 0$ such that, for $n$ large enough

$$P^\pi \left[ \{ f \in F^+: l(f, f_0) > C n^{-2\beta/(2\beta+1)} \log n^{(2\beta+3)/(2\beta+1)} \} | X_n \right] \leq C' n^2$$

and

$$E_0^\pi \left[ l(\hat{f}, f_0) \right] \leq 2 C n^{-2\beta/(2\beta+1)} \log n^{(2\beta+3)/(2\beta+1)},$$

where $\log \hat{f}(\lambda) = E^\pi [\log f(\lambda) | X_n]$.

**Proof.** Throughout the proof, $C$ denotes a generic constant. The proof of the theorem is divided in two parts; in the first part, we prove that

$$E_0^\pi \left[ P^\pi \left\{ f : h_n(f, f_0) \geq n^{-2\beta/(2\beta+1)} \log n^{(2\beta+3)/(2\beta+1)} \right\} | X_n \right] \leq \frac{C}{n^2}$$

and in the second part we prove that

$$h_n(f, f_0) \leq C n^{-2\beta/(2\beta+1)} \log n^{1/\beta} \Rightarrow l(f, f_0) \leq C' n^{-2\beta/(2\beta+1)} \log n^{(2\beta+3)/(2\beta+1)},$$

(11)
for some constant $C' > 0$, when $n$ is large enough. The latter inequality implies that

$$E_n \left[ l(f, f_0) | X_n \right] \leq C' n^{-\frac{2\beta}{2\beta+1}} \log n^{\frac{2\beta+4}{2\beta+1}} + \int_{A(n, \beta)} l(f, f_0) d\pi(f | X_n) \leq 2C' n^{-\frac{2\beta}{2\beta+1}} \log n^{\frac{2\beta+4}{2\beta+1}},$$

for large $n$, where $A(n, \beta) = \{ h_n(f, f_0) > Cn^{-\frac{2\beta}{2\beta+1}} \log n^{\frac{2\beta+3}{2\beta+1}} \}$. This would imply Theorem 3.

To prove (10), we need to show that conditions (i)-(iii) of Theorem 2 are fulfilled. Condition (ii) is obvious because the prior has the same form as in Section 2.3 and, because when $f \in \mathcal{S}(\beta, L)$, there exists $A > 0$ such that $\sum_j |\theta_j| \leq A$. Let $\epsilon_n^2 = n^{-2\beta/(2\beta+1)} \log n^{(2\beta+3)/(2\beta+1)}$, let $K_n = k_0 n^{1/(2\beta+1)} \log n^{2/(2\beta+1)}$, $d \leq d_0 \leq d + \epsilon_n / \log n^{3/2}$ and, for all $l = 0, \ldots, K_n$, $|\theta_l - \theta_{0l}| \leq (l + 1)^{-\beta/2}(\log (l + 1))^{-1/2}$, where $f_0 \in \mathcal{S}(\beta, L_0)$, $\exists t_0 > 0$ such that

$$\sum_{l \geq K_n} \theta_{0l}^2 \leq L_0 K_n^{-2\beta} \leq C_n^2 (\log n)^{-3}, \quad \sum_{l \geq K_n} |\theta_{0l}| \leq K_n^{-t_0}$$


We now show that assumption (i) of Theorem 2 is satisfied. Since

$$KL_n(f_0; f) \leq h_n(f, f_0) = \frac{1}{2n} \text{tr} \left( T_n(f_0 - f) T_n^{-1}(f) T_n(f_0 - f) T_n^{-1}(f_0) \right),$$

it will be enough to prove the assumption under the above conditions for $h_n(f, f_0) \leq C\epsilon_n^2$.

Let $f_{0n}(\lambda) = |1 - e^{i\lambda}|^{-2d_0} \exp \left( \sum_{l=0}^{K_n} \theta_{0l} \cos l \lambda \right)$, $h_n(\lambda) = 1 - \exp \left( - \sum_{l \geq K_n+1} \theta_{0l} \cos l \lambda \right)$, and $g_n = f_{0n}^{-1}(f_{0n} - f)$; then $f_0 - f = f_0 b_n + f_0 g_n$ and

$$nh_n(f_0, f) \leq \text{tr} \left( T_n(f_0 b_n) T_n^{-1}(f) T_n(f_0 b_n) T_n^{-1}(f) \right) + \text{tr} \left( T_n(f_0 g_n) T_n^{-1}(f) T_n(f_0 g_n) T_n^{-1}(f) \right).$$


Both terms of the right hand side of (13) are treated similarly using equation (25) of Lemma 3, that is
\[
\text{tr} \left( T_n(f_0 b_n) T_n^{-1} (f) T_n(f_0 b_n) T_n^{-1} (f_0) \right) \leq C (\log n)^3 n |b_n|^{1/2} + O(n^\delta)
\]
and
\[
\text{tr} \left( T_n(f_0 g_n) T_n^{-1} (f) T_n(f_0 g_n) T_n^{-1} (f_0) \right) \leq C \text{tr} \left( T_n(f_0 g_n) T_n^{-1} (f) T_n(f_0 g_n) T_n^{-1} (f_0) \right) \leq C (\log n)^3 n |g_n|^{1/2} + O(n^\delta).
\]
This implies that
\[
h_n(f_0, f) \leq C \epsilon^2 n, \quad \text{when } f \text{ satisfies the conditions described above and}
\]
\[
B_n \subset \left\{ f_{k,d,\theta}; k \geq K_n, d \leq d_0 \leq d + \frac{\epsilon_n}{\log n}, 0 \leq l \leq K_n, |\theta_l - \theta_{0l}| \leq \frac{(l + 1)^{-\beta} \epsilon_n}{(\log (l + 1)) \log n^{3/2}} \right\}.
\]
The prior probability of the above set is bounded from below by
\[
\pi(K_n) \mu_1 \left( (\eta_1, \ldots, \eta_K_n) : |\eta_l - \eta_{0l}| \leq C \frac{1}{\log l} \epsilon_n \log n^{-3/2} \right),
\]
where \(\mu_1\) denotes the uniform measure on the set \(\{ (\eta_1, \ldots, \eta_K_n); \sum \eta_l^2 \leq A \}\). We finally obtain that
\[
\pi(B_n(\delta)) \geq e^{-CK_n \log n} \geq e^{-n \rho_n / 2}
\]
by choosing \(k_0\) small enough, and condition (i) of Theorem 3 is satisfied by the FEXP(\(\beta\)) prior. We now verify condition (iii) of Theorem 3. Let \(\mathcal{F}_n = \{ f_{\theta,k}; k \leq K_n \}\), with \(K_n = K_0 n^{1/(2\beta+1)} \log n^{2/(2\beta+1)}\), let \(j_0 \leq j \leq J_n\), where \(j_0\) is some positive constant, and consider \(f \in S_{n,j}\), as defined in Theorem 2, where \(f(\lambda) = f_{\theta,k} = |1 - e^{i\lambda}|^{-2d} \exp\{ \sum_{l=1}^k \theta_l \cos (l\lambda) \}\). Define
\[
f_u(\lambda) = |1 - e^{i\lambda}|^{-2d - c \epsilon_n^2} \exp\{ \sum_{l=1}^k \theta_l \cos (l\lambda) + c \epsilon_n^2 \},
\]
for some constant \(c > 0\). Then if \(f' \) is such that
\[
|d - d'| \leq c \rho_n t / 2, \quad |\theta'_l - \theta_l| \leq (l_j + 1)^{-\beta} \log (l + 1)^{-1} \epsilon_n^2 / 2,
\]
\[ 0 \leq (f_u - f') (\lambda) \leq 4c \epsilon_n^2 j \left( (\log \lambda)^2 + 1 \right) f_u (\lambda), \quad f' (\lambda) \geq e^{-2c \epsilon_n^2 j} f_u (\lambda) \delta_n (\lambda), \]

where \( \delta_n (\lambda) = (1 - \cos (1 - \lambda))^{-2c \epsilon_n^2 j} \) and

\[
\text{tr} \left( T_n^{-1} (f') T_n (f_u - f') \right) \leq 4c \epsilon_n^2 j e^{2c \epsilon_n^2 j} \text{tr} \left( T_n^{-1} (f_u \delta_n) T_n (f_u) \right) \\
\leq C c \epsilon_n^2 j \\
\leq C c_h n (f_0, f_u).
\]

By choosing \( c \) small enough we obtain that \( \text{tr} \left( T_n^{-1} (f') T_n (f_u - f') \right) \leq n h_n (f_0, f_u) / 8 \). Similarly

\[
\text{tr} \left( T_n^{-1} (f_0) T_n (f_u - f') \right) \leq 4c \epsilon_n^2 j \text{tr} \left( T_n^{-1} (f_0) T_n (f_u) \right) \\
\leq C c h_n (f_0, f_u) / 8.
\]

Since we are in the set \( \{ f; d(f_0, f) \leq \epsilon \} \), for some \( \epsilon > 0 \) fixed but as small as we need, there exists \( \epsilon', \epsilon'' > 0 \) such that

\[
|d - d_0| < \epsilon', \quad \sum_{i=1}^K (\theta_i - \theta_0)^2 + \sum_{t \geq K+1} \theta_{t0}^2 \leq \epsilon''.
\]

Let \( K \leq K_n = K_0 n^{1/(2\beta+1)} (\log n)^{-1} \), the number of \( f_u \) defined as above in the set \( S_{n,j} \) is bounded by

\[
N_{n,j} \leq K_n (1 - \epsilon_n^2) (C K_n j^{-1} \epsilon_n^2)^K_n
\]

and

\[
\tilde{N}_{n,j} = \log N_{n,j} \leq c_j \epsilon_n^2
\]

where \( c_j \) is decreasing in \( j \). Hence by choosing \( j_0 \) large enough condition (iii) is verified by the \( \text{FEXP}(\beta) \) prior. This achieves the proof of (10) and we obtain a rate of convergence, in terms of the distance \( h_n (., .) \). We now prove (11) to obtain a rate of convergence in terms of the distance \( l(., .) \). Consider \( f \) such that

\[
h_n (f_0, f) = \frac{1}{n} \text{tr} \left( T_n^{-1} (f_0) T_n (f - f_0) T_n^{-1} (f) T_n (f - f_0) \right) \leq \epsilon_n^2.
\]
Equation (26) of Lemma 3 implies that
\[
\frac{1}{n} \text{tr} \left( T_n(f_0^{-1})T_n(f - f_0)T_n(f^{-1})T_n(f - f_0) \right) \leq C\epsilon_n^2,
\]
leading to
\[
\frac{1}{n} \text{tr} \left( T_n(g_0)T_n(f - f_0)T_n(g)T_n(f - f_0) \right) \leq C\epsilon_n^2, \tag{14}
\]
where \( g_0 = (1 - \cos \lambda)^d_0, \ g = (1 - \cos \lambda)^d. \)

We now prove that \( \text{tr} \left( T_n(g_0(f - f_0))T_n(g(f - f_0)) \right) \leq C\epsilon_n^2; \) Using the same calculations as in the control of \( I_2 \) in Appendix C,
\[
\Delta = \frac{1}{n} \text{tr} \left( T_n(g_0(f - f_0))T_n(g(f - f_0)) \right) - \frac{1}{n} \text{tr} \left( T_n(g_0)T_n(f - f_0)T_n(g)T_n(f - f_0) \right)
\]
\[= O(n^{\delta-1} \log n^3),\]
as soon as \( |d - d_0| < \delta/2, \) where \( \delta \) is any positive constant, as small as we want. This implies, together with (14) that
\[
\frac{1}{n} \text{tr} \left( T_n(g_0(f - f_0))T_n(g(f - f_0)) \right) \leq C\epsilon_n^2.
\]

To finally obtain (11), we use equation (27) in Lemma 3 which implies that
\[
A_n = \text{tr} \left( T_n(g_0(f - f_0))T_n(g(f - f_0)) \right) - \text{tr} \left( T_n(g_0g(f - f_0)^2) \right)
\]
\[\leq Cn^{-1+\delta} + \log n \sum_{l=0}^{K_n} |\theta_l| \left( \int_{[-\pi,\pi]} g_0 g(f - f_0)^2(\lambda) d\lambda \right)^{1/2}.\]

Moreover
\[
\sum_{l=1}^{K_n} |\theta_l| \leq \sum_{l=1}^{l^{2\beta+r}} \theta^2 + \sum_{l=1}^{l^{-r/(2\beta-1)}} l^{1-r/(2\beta-1)} \leq CK_n^r + K_n^{1-r/(2\beta-1)},
\]
by choosing \( r = (2\beta - 1)/2\beta \), \( A_r/n \) is of order \( n^{-(4\beta^2+1)/(2\beta(2\beta+1))} \) which is negligible compared to \( n^{-2\beta/(2\beta+2)} \) so that if \( \beta \geq 1/2 \)

\[
\int_{[-\pi,\pi]} g_0 g(f_0 - f)^2 d\lambda \leq \epsilon_n^2,
\]

which achieves the proof. ■

4 Computational issues

Any practical implementation of our nonparametric approach must take into account the fact that the computation of the likelihood function in this context is very expensive since both the determinant and the inverse of a large Toeplitz matrix must be computed at each evaluation. Then one should prefer to use a Monte Carlo approximation which is as easiest as possible to handle and with fast convergence properties. We consider several different approaches, which are discussed in a companion paper (Liseo and Rousseau 2006). Our final preference was for a modification of the adaptive Monte Carlo algorithm, proposed and discussed in Douc et al. (2006), which can also be used in the presence of variable dimension parametric spaces as is the case here. The main advantage of adaptive Monte Carlo algorithms is that they do not rely upon asymptotic convergence of the samplers but, rather, they should be considered as an evolution of the importance sampling schemes, with the advantage of an on-line adaptation of the weights given to the proposal distributions, which are allowed to be more than one. Here we briefly describe the main features of the proposed algorithm. Further details can be found in Liseo and Rousseau (2006). Denote by \((K, \eta_K)\) our global parameter, with \( \eta_K = (\theta_0, \ldots, \theta_k, d) \); here \( K \) is a positive integer which determines the size of the parameter vector. First one has to define a set of proposal distributions, which we denote by \( Q_h(\cdot, \cdot), h = 1, \ldots, H \); they represent \( H \) possible different kernels, which we assume are all dominated by the same dominating measure, as the poste-
rior distribution $\pi_x$. Denote by $q_h(\cdot, \cdot)$ the corresponding densities. Then one has to perform $T$ different iterations of an importance sampling approximation, each based on $N$ proposed values. The novel feature is that, at each $t$, $t = 1, \cdots, T$, the weights of the sampled values are calibrated in terms of the previous iterations. From a practical perspective, one should work with a large value of $N$, and a small value (say 5-10) of $T$. Douc et al. (2006) show that, in fix dimension problems, the algorithm will converge toward the optimal mixture of proposals, in few iterations, at least in a Kullback-Leibler distance sense.

The algorithm follows quite closely the one described in Douc et al. (2006) and will not reported here. The only significant difference is that we have to deal with the variable dimension of the parameter space; then we must be able to propose a set of possible “moves” to subspaces with a different value of $k$. Then, at each iteration $t$, ($t = 1, \cdots, T$) and for each sample point $j$, ($1 \leq j \leq N$), we propose a new value $K'_{j,t}$ from a distribution on the set of integers, conditional on the previous value of $K_j$; then, conditionally on $K'_{j,t}$ and on the value of $\eta_{j,t-1}$, propose a new value $\eta_{j,K'_{j,t}}^{(t)}$. The description of all the possible moves is quite involved; here we sketch the ideas behind the strategy. For a fixed $1 \leq j \leq N$, at the $t$-th iteration of the algorithm, one draws a new value $K'_{j,t}$, according the following proposals for $K'$:

$$K_{j,t} = K_{j,t-1} + \xi_{j,t}, \quad \text{where} \quad \xi_{j,t}|K_{j,t-1} \sim p_1 \text{Po}(\lambda_1) + (1 - p_1) \text{NePo}_{K_{j,t-1}}(\lambda_2),$$

where $p_1 \in (0, 1)$ and the symbol NePo$_k$ denotes a truncated Poisson distribution over the set \{-$k$, -$k+1$, $\ldots$, 0\}. At each iteration the proposed value $K_{j,t}$ may be either less than, equal to, or larger than $K_{j,t-1}$:

- If $K_{j,t} < K_{j,t-1}$ then $\theta_{K_{j,t+1}}^{(t)} = \cdots = \theta_{K_{j,t-1}}^{(t)} = 0$ and

\begin{equation}
(\theta_0^{(t)}, \ldots, \theta_{K_{j,t}}^{(t)}) = (\theta_0^{(t-1)}, \ldots, \theta_{K_{j,t-1}}^{(t-1)}) + \varepsilon_{K_{j,t}}
\end{equation}

(15)
and $\varepsilon_{K_{j,t}}$ is a $K_{j,t}$-dimensional symmetric proposal.

- If $K_{j,t} = K_{j,t-1}$ then the new point is draw according to (15), without eliminating any parameter.
- If $K_{j,t} > K_{j,t-1}$, then the first $K_{j,t-1}$ are drawn according to (15), while the latest components are drawn from the same kernel proposal $\varepsilon$, although centered on an easy-to-calculate point estimate obtained from a simplified version of the estimator presented in Hurvich et al. (2002). Notice that, within this approach, even when the algorithm proposes a change of dimension, one must propose a global move for all the components of the parameter vector, in order to satisfy the positivity condition on the proposal required by Douc et al. (2006). For the practical evaluation of the likelihood function, we have used the approximations of the inverse matrix and of the determinant of a Toeplitz matrix, as proposed in Chen, Hurvich and Lu (2006).

4.1 Analysis of Nile river data

Here we analyze the Nile river data that consists of the time series of annual minimum water levels of the River Nile at the Roda Gorge; the data are available, for example, in Beran 1994, page 237. We examined $n = 512$ observations corresponding to the period from 622 to 1133. A visual study (see Figure 1) of the data and the ACF reveals a strong persistence of autocorrelations; after removing the mean, we applied our procedure to produce the corresponding spectral estimate; a point-wise 95% credible band is computed from the PMC sample and is plotted in Figure 2. Figure 3 reports the PMC histogram approximation to the posterior distribution of $d$. Notice that the posterior mean estimate of $d$ is in agreement with other proposed estimates: see, for example, Robinson (1995).
Figure 1: Nile river time series data and Autocorrelation function
Figure 2: Spectral density for Nile data
Figure 3:
Appendices

A Lemmas 1 and 2

We state two technical lemmas, which are extensions of Lieberman, Rousseau and Zucker (2003) on uniform convergence of traces of Toeplitz matrices, and which are repeatedly used in the paper.

**Lemma 1** Let $t > 0$, $M > 0$ and $\bar{M}$ a positive function on $]0, \pi[$, let $p$ be a positive integer, and

\[ F(d, M, \bar{M}) = \left\{ f \in \tilde{F}, \forall u > 0, \sup_{|\lambda| > u} \frac{d\tilde{f}(\lambda)}{d\lambda} \leq \bar{M}(u) \right\}, \]

we have:

\[ \sup_{p(d_1 + d_2) \leq 1/2 - t} \left| \frac{1}{n} \text{tr} \left( \prod_{i=1}^{p} T_n(f_i)T_n(g_i) \right) - (2\pi)^{2p-1} \int_{-\pi}^{\pi} \prod_{i=1}^{p} f_i(\lambda)g_i(\lambda)d\lambda \right| \to 0. \quad (16) \]

and let $L > 0$ and $\rho \in (0, 1]$

\[ \sup_{p(d_1 + d_2) \leq 1/2 - t} \left| \frac{1}{n} \text{tr} \left( \prod_{i=1}^{p} T_n(f_i)T_n(g_i) \right) - (2\pi)^{2p-1} \int_{-\pi}^{\pi} \prod_{i=1}^{p} f_i(\lambda)g_i(\lambda)d\lambda \right| \to 0. \quad (17) \]

This lemma is an obvious adaptation from Lieberman *et al.* (2003), and the only non obvious part is the change from the condition of continuous differentiability in that paper to the Lipschitz condition of order $\rho$, considered equation 17. This different assumption affects only equation (30) of Lieberman *et al.* (2003), with $\eta_n$ replaced by $\eta_n', \rho_1$ which does not change the convergence results.

**Lemma 2**

\[ \sup_{2p(d_1 - d_2) \leq 1/2 - t} \left| \frac{1}{n} \text{tr} \left( \prod_{i=1}^{p} T_n(f_i)T_n(g_i) \right) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{i=1}^{p} \frac{f_i(\lambda)}{g_i(\lambda)}d\lambda \right| \to 0, \]
\[
\sup_{d_1 - d_2 \leq \rho^2 \wedge 1/2 - t}
\left| \frac{1}{n} \tr \left( \prod_{i=1}^{p} T_n(f_i) T_n(g_i)^{-1} \right) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{i=1}^{p} \frac{f_i(\lambda)}{g_i(\lambda)} d\lambda \right| \to 0.
\]

and

\[
\sup_{d_1 - d_2 \leq \rho^2 \wedge 1/2 - t}
\left| \frac{1}{n} \tr \left( \prod_{i=1}^{p} T_n(f_i) T_n(g_i)^{-1} \right) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{i=1}^{p} \frac{f_i(\lambda)}{g_i(\lambda)} d\lambda \right| \to 0.
\]

**Proof.** Proof of lemma 2.

In this second lemma, the uniformity result is a consequence of the first lemma, as in Lieberman et al. (2003); The only difference is in the proof of Lemma 5.2. of Dahlhaus (1989), i.e. in the study of terms in the form

\[ |\text{id} - T_n(g)^{1/2} T_n((4\pi^2 g)^{-1}) T_n(g)^{1/2}|. \]

Following Dahlhaus (1989)'s proof, we obtain an upper bound of

\[ \frac{|g(\lambda_1)|}{g(\lambda_2)} - 1 \]

which is different from Dahlhaus (1989)'s. If \( g \in G(d_2, m, M, L, \rho_2) \), the Lipschitz condition in \( \rho \) implies that

\[ \frac{|g(x)|}{g(y)} - 1 \leq K \left( |x - y|^\rho + \frac{|x - y|^{1-\delta}}{|x|^{1-\delta}} \right). \]

Calculations using \( L_N \) as in Dahlhaus (1989) imply that

\[ |I - T_n(f)^{1/2} T_n((4\pi^2 f)^{-1}) T_n(f)^{1/2}|^2 = O(n^{1-2\rho \log n}) + O(n^\delta), \quad \forall \delta > 0. \]

If \( g \in L^*(M, m, L) \) as defined in Section 3.2, then

\[ \left| \frac{f(x)}{f(y)} - 1 \right| \leq K \left( \frac{|x - y|^{1-3\delta}}{(|x| \wedge |y|)^{1-\delta}} \right) \leq K |x - y|^{1-3\delta} \left( \frac{1}{|x|^{1-\delta}} + \frac{1}{|y|^{1-\delta}} \right) \]
and Dahlhaus (1989) Lemma 5.2 is proved, leading to a constraint in the form $4p(d_1 - d_2) < 1$ (corresponding to $\rho = 1$).

Then, using again Dahlhaus (1989)’s calculations we obtain that

$$|A - B| = 0(n^{2(d_2 - d_1)} n^{1/2 - (\rho \wedge 1/2) + \delta}), \quad \forall \delta > 0$$

and finally that

$$\frac{1}{n} \text{tr} \left\{ \prod_{j=1}^{p} A_j - \prod_{j=1}^{p} B_j \right\} = \sum_{k=1}^{p} O(n^{-1/2} n^{2(p-k)(d_2 - d_1)} n^{2(d_2 - d_1) n^{1/2 - \rho}})$$

which goes to zero when $2p(d_2 - d_1) < \rho \wedge 1/2$. ■

**B Proof of Theorem 1**

Before giving the proof of Theorem 1, we give a more general version of assumption (ii), namely (ii)bis, ensuring consistency of the posterior. It is quite cumbersome in its formulation, but we believe that it might prove useful in some context.

**Assumption [(ii)bis]** For all $\varepsilon > 0$ there exists $F_n \subset \{ f \in \bar{\mathcal{F}}_+, d(f_0, f_i) > \varepsilon \}$, such that $\pi(F_n^c) \leq e^{-nr}$ and there exist $t, M, m > 0$ with $t < \rho_0/4$, and a smallest possible net $\mathcal{H}_n \subset \{ \bar{\mathcal{F}}_+(t, M, m), d(f, f_0) > \varepsilon/4 \}$ such that $\forall f \in F_n, \quad \exists f_i \geq f \in \mathcal{H}_n$ satisfying either of the three conditions

1. $|4(d_0 - d_i)| \leq \rho_0 \wedge 1/2 - t$ and
   $$\max \left( \frac{\text{tr}[T_n(f)^{-1}T_n(f_i) - \text{id}]}{n}, \frac{\text{tr}[T_n(f_i - f)T_n^{-1}(f_0)]}{n} \right) \leq \frac{h_n(f_0, f_i)}{8}. \quad (18)$$

2. $4(d_i - d_0) > \rho_0 \wedge 1/2 - t$
   $$\text{tr}[T_n(f)^{-1}T_n(f_i) - \text{id}]/n \leq KL_n(f_0; f_i)/4, \quad h_n(f, f_i) \leq KL_n(f_0; f_i)/4, \quad (19)$$
3. $4(d_0 - d_i) > \rho_0 \wedge 1/2 - t$ and

$$\frac{1}{n} [T_n(f)T_n^{-1}(f_i) - \text{id} - T_n(f - f_i)T_n^{-1}(f_0)] \leq KL_n(f_i; f_0)/2$$

(20)

We also assume that $\forall f \in \mathcal{H}_n$,

$$\frac{1}{n} \text{tr} \left((T_n(|\lambda|^{-2d})T_n(f_0 - f))^2\right) \geq B_1B(f, f_0) \quad \text{if} \quad d \leq d_0$$

and

$$\frac{1}{n} \text{tr} \left((T_n(|\lambda|^{-2d_0})T_n(f_0 - f))^2\right) \geq B_1B(f_0, f) \quad \text{if} \quad d \geq d_0$$

**Proof of Theorem 1**

The proof follows the same ideas as in Ghosal *et al.* (2000). We can write

$$P_\pi[A_\epsilon|X_n] = \frac{\int_{A_\epsilon} \varphi_f(X_n)/\varphi_{f_0}(X_n)d\pi(f)}{\int_{\mathcal{F}_0} \varphi_f(X)/\varphi_{f_0}(X_n)d\pi(f)} = \frac{N_n}{D_n}.$$ 

Then the idea is to bound from below the denominator using condition (i) of the Theorem and to bound from above the numerator using a discretization of $A_\epsilon$ based on the net $\mathcal{H}_n$ defined in (ii) of the Theorem and on tests.

Let $\delta, \delta_1 > 0$: one has

$$P_0 \left[P_\pi[A_\epsilon|X_n] \geq e^{-n\delta} \right] \leq P_0^n \left[D_n \leq e^{-n\delta_1} \right] + P_0^n \left[N_n \geq e^{-n(\delta + \delta_1)} \right]$$

(21)

Also, let

$$\mathcal{B}_n = \{f : \text{tr} \left((B(f_0, f)) - \log \det(A(f_0, f)) \right) \leq n\delta_1; \text{tr} \left(B(f_0, f)^3\right) \leq M'n\},$$

where

- $A(f_0, f) = T_n(f)^{-1}T_n(f_0)$
\[
\begin{align*}
B(f_0, f) &= T_n(f_0)^{1/2}[T_n(f)^{-1} - T_n(f_0)^{-1}]T_n(f_0)^{1/2}, \\
\text{and } M > 0 \text{ is fixed, and define } \\
\Omega_n &= \{(f, X) : -X^t[T_n(f)^{-1} - T_n(f_0)^{-1}]X + \log(\det(A_n)) > -2n\delta_1\}.
\end{align*}
\]

We then have
\[
\begin{align*}
p_1 &\leq P_0^n \left( \int_{\Omega_n \cap B_n} \frac{\varphi_f(X)}{\varphi_{f_0}(X)} d\pi(f) \leq e^{-n\delta_1/2} \frac{\pi(B_n)}{2} \right) \\
&\leq P_0^n \left( \pi(B_n \cap \Omega_n) \leq \frac{\pi(B_n)}{2} \right) \\
&\leq P_0^n \left( \pi(B_n \cap \Omega_n^c) > \frac{\pi(B_n)}{2} \right) \\
&\leq 2 \int_{B_n} P_0^n[\Omega_n^c]|d\pi(f)| \frac{\pi(B_n)}{\pi(B_n)}.
\end{align*}
\]

Moreover,
\[
\begin{align*}
P_0^n[\Omega_n^c] &= P_0^n \left( X_n^t[T_n(f)^{-1} - T_n(f_0)^{-1}]X_n + \log(\det(A(f_0, f))) > 2n\delta_1 \right) \\
&= Pr[y^t B(f_0, f)y - tr(B(f_0, f)) > 2n\delta_1 + \log(\det(A(f_0, f))) - tr(B(f_0, f))]
\end{align*}
\]

where \( y \sim N_n(0, \text{id}) \). When \( f \in B_n, 2n\delta_1 + \log(\det(A(f_0, f))) - tr(() B(f_0, f)) > n\delta_1, \) so that
\[
\begin{align*}
P_0^n[\Omega_n^c] &\leq Pr[y^t B(f_0, f)y - tr(() B(f_0, f)) > n\delta_1] \\
&\leq \frac{E[(y^t B(f_0, f)y - tr(() B(f_0, f)))^3]}{n^3}.
\end{align*}
\]

Moreover,
\[
E[(y^t B(f_0, f)y - tr(() B(f_0, f)))^3] = 8 tr(() B(f_0, f))^3 \leq M'n,
\]

whenever \( f \in B_n \). Hence, \( p_1 \leq 8M'/n^2 \). Besides,
\[
B_n \subset \{ f \in \tilde{G}(t, M, L, \rho) ; KL_{\infty}(f_0; f) \leq \frac{\delta_1}{2\delta_5} 6(d_0 - d) \leq \rho - t, \frac{1}{2\pi} \int (\frac{f_0}{f} - 1)^3 dx \leq \frac{M'}{2} \}.
\]
so that assumption (i) implies that, for $n$ large enough, $\pi(B_n) \geq e^{-n\delta_1}/2$ and

$$P_0^n \left[ D_n \leq e^{-n\delta_1} \right] \leq \frac{8M'}{n^2}.$$ 

We now consider the second term of (21), namely:

$$p_2 = P_0^n \left[ N_n \geq e^{-n(\delta+\delta_1)} \right] \leq 2e^{n(\delta+\delta_1)} \pi(F_n^c) + P_0^n \left[ \int_{A_n \cap F_n} \varphi f(X_n) d\pi(f) \geq e^{-n(\delta+\delta_1)}/2 \right] \leq e^{-n(r-(\delta+\delta_1))} + \tilde{p}_2,$$

take $\delta + \delta_1 < r$ and consider $\tilde{p}_2$. Consider the following tests: let $f_i \in H_n$,

$$\phi_i = \mathbb{1}_{X'(T_n^{-1}(f_0) - T_n^{-1}(f_i))X \geq n \rho_i},$$

1. If $4(d_0 - d_i) \leq \rho_0 - t$ and $4(d_i - d_0) \leq \rho_0 - t$, then if

$$\rho_i = \operatorname{tr} \left( \mathbf{id} - T_n(f_0)T_n^{-1}(f_i) \right)/n + h_n(f_0, f_i)/4,$$

$$E_0^n [\phi_i] \leq \max \left( \exp \left\{ -n \frac{h_n(f_0, f_i)^2}{512b_i} \right\}, \exp \left\{ -n \frac{h_n(f_0, f_i)^2}{32} \right\} \right),$$

(22)

where $b_i = n^{-1} \operatorname{tr} \left( (\mathbf{id} - T_n(f_0)T_n^{-1}(f_i))^2 \right)$, and for any $f \in F_+(1-t, M, m)$ satisfying $f \leq f_i$ and

$$\frac{1}{n} \operatorname{tr} \left( T_n(f)T_n^{-1}(f_i) - \mathbf{id} \right) + \frac{1}{n} \operatorname{tr} \left( T_n(f_i - f)T_n^{-1}(f_0) \right) \leq h_n(f_0, f_i)/4,$$

we have

$$E_0^n [1 - \phi_i] \leq \max \left( \exp \left\{ -n \frac{h_n(f_0, f_i)^2}{16B_i} \right\}, \exp \left\{ -n \frac{h_n(f_0, f_i)}{4} \right\} \right),$$

where $B_i = n^{-1} \operatorname{tr} \left( (\mathbf{id} - T_n(f_i)T_n^{-1}(f_0))^2 \right)$. 

36
2. If $4(d_i - d_0) > \rho_0 - 4t$, if $\rho_i = \text{tr} \left( \text{id} - T_n(f_0)T_n^{-1}(f_i) \right) / n + KL_n(f_0; f_i)/2$,
\[
E_{\rho_i} \leq \max \left( \exp \left\{ -n \frac{KL_n(f_0; f_i)^2}{512b_i} \right\}, \exp \left\{ -n \frac{h_n(f_0, f_i)}{32} \right\} \right),
\]
and for all $f \leq f_i$ satisfying
\[
\text{tr} \left( T_n(f)^{-1}T_n(f_i) - \text{id} \right) \leq nKL_n(f_0; f_i)/4, \quad \text{and} \quad h_n(f, f_i) \leq KL_n(f_0; f_i)/4,
\]
we have
\[
E_{\rho_i} \leq e^{-nKL_n(f_0; f_i)/2}
\]

3. If $4(d_0 - d_i) > \rho_0 - 4t$, if $\rho_i = \log \det[T_n(f_i)T_n(f_0)^{-1}]/n$, then
\[
E_{\rho_i} \leq \max \left( \exp \left\{ -n \frac{KL_n(f_i; f_0)^2}{8B_i} \right\}, \exp \left\{ -n \frac{KL_n(f_i; f_0)}{4} \right\} \right).
\]
Moreover, for all $f \in G(1 - t, M_1, M_2)$, satisfying $f \leq f_i$ and
\[
\frac{1}{n} \text{tr} \left( T_n(f)T_n^{-1}(f_i) - \text{id} - T_n(f - f_i)T_n^{-1}(f_0) \right) \leq KL_n(f_i; f_0),
\]
we have
\[
E_{\rho_i} \leq \max \left( \exp \left\{ -n \frac{KL_n(f_i; f_0)^2}{64B_i} \right\}, \exp \left\{ -n \frac{KL_n(f_i; f_0)}{8} \right\} \right).
\]

The difficulty is now to transform these conditions into a net. Using Dahlhaus’s type of calculation (Dahlhaus (1989), pag. 1755) there exists a constant $C > 0$ (depending on $M, m$) uniformly over the class $\cup_{d \leq d_0} F_+(d, m, M)$, such that
\[
KL_n(f_i; f_0) \geq \frac{1}{4C^2} \frac{1}{n} \text{tr} \left( (T_n(f_0)^{-1}T_n(f_i) - \text{id})^2 \right) \quad (23)
\]
for all $n \geq 1$. Similarly,
\[
KL_n(f_0; f_i) \geq \frac{1}{4C^2} \frac{1}{n} \text{tr} \left( (T_n(f_i)^{-1}T_n(f_0) - \text{id})^2 \right) \quad (24)
\]
Since for all $f \in \mathcal{H}_n$, associated with the long-memory parameter $d$,
\[
\frac{1}{n} \text{tr} \left( (T_n(|\lambda|^{-2d})T_n(f_0 - f))^2 \right) \geq B_1 B(f, f_0) \quad \text{if} \quad d \leq d_0
\]
and
\[
\frac{1}{n} \text{tr} \left( (T_n(|\lambda|^{-2d_0})T_n(f_0 - f))^2 \right) \geq B_1 B(f_0, f) \quad \text{if} \quad d \geq d_0
\]

We end up with an upper bound in the form:
\[
E_n^n [\phi_i] \leq e^{-nc B(f_i, f_0)}, \quad e^{-nc B(f_0, f_i)}
\]
\[
E_f^n [1 - \phi_i] \leq e^{-nc B(f_i, f_0)}, \quad e^{-nc B(f_0, f_i)}
\]

When $d(f_i, f_0) > \varepsilon$, up to the $(2\pi)^{-1}$ term
\[
B(f_i, f_0) = \int \left( \frac{f_i}{f_0} - 1 \right)^2 dx
\]
\[
\geq c \frac{KL_\infty(f_i; f_0)}{|\log KL_\infty(f_i; f_0)|}
\]
\[
\geq c \varepsilon |\log \varepsilon|^{-1},
\]
for some constant $c$. The last inequalities comes from the fact that $(f_i/f_0 - 1 - \log f_i/f_0) \leq C(f_i/f_0 - 1)^2$ unless $f_i/f_0$ is close to zero ($A = \{f_i/f_0 \leq x_c\}$). On this set, we bound
\[
\int_A \frac{f_i}{f_0} - 1 - \log f_i/f_0)(\lambda)d\lambda \leq \int_A \log f_0/f_i(\lambda)d\lambda
\]
\[
\leq (\int_A d\lambda) \log \left( \frac{\int_A f_0/f_i(\lambda)d\lambda}{\int_A d\lambda} \right)
\]
\[
\leq C'(\int_A d\lambda) \log(\int_A d\lambda),
\]
for some constant $C' > 0$.

Finally, in each case, we have, when $n$ is large enough (independently of $f_i$),
\[
E_0^n [\phi_i] \leq e^{-nc |\log \varepsilon|^{-1}} \leq e^{-nc |\log \varepsilon|^{-2}}
\]
for all $\varepsilon < \varepsilon_0$, and any $\delta > 0$, and

$$E_f^n [1 - \phi_i] \leq e^{-n\varepsilon|\log \varepsilon|^{-2}}.$$ 

Let $\phi^n = \max_i \phi_i$, we then have that

$$\hat{p}_2 \leq E_0^n [\phi^n] + \int_{A_i \cap F_n} E_f [1 - \phi^n] d\pi(f) \leq e^{N_n} e^{-n\varepsilon|\log \varepsilon|^{-2}} + e^{-n\varepsilon|\log \varepsilon|^{-2}} \leq 2e^{-n\varepsilon|\log \varepsilon|^{-2}/2}$$

and the theorem follows under assumption (ii)bis. To obtain assumption (ii) on the net on $F_n$ given in Theorem 1 we have $0 \leq f_i(\lambda) - f(\lambda) \leq cf_i(\lambda)|\lambda|^{-t/2}\varepsilon|\log \varepsilon|^{-1}$, then using the same kind of calculations as previously we have that

$$\frac{1}{n} \text{tr} \left( T_n^{-1}(f)T_n(f_i - f) \right) \leq \frac{Mc\varepsilon|\log \varepsilon|^{-1}}{nm} \text{tr} \left( T_n^{-1}(|\lambda|^{-2d})T_n(|\lambda|^{-2d_i-t/2}) \right) \leq 2\frac{Mc\varepsilon}{m} \int |\lambda|^{2(d-d_i)-t/2} d\lambda,$$

where the latter inequality comes from Lemma 2. Similarly

$$\frac{1}{n} \text{tr} \left( T_n^{-1}(f_0)T_n(f_i - f) \right) \leq 2\frac{Mc\varepsilon|\log \varepsilon|^{-1}}{m} \int |\lambda|^{2(d_0-d_i)-t/2} d\lambda,$$

Then as seen previously

$$KL_n(f_0; f_i) \geq C_1 B(f_0, f_i), \quad KL_n(f_i; f_0) \geq C_1 B(f_i, f_0)$$

depending on the sign of $d_0 - d_i$, with $C_1$ some fixed positive constant, so that by choosing $c$ small enough the conditions required for the net are satisfied by the above $f_i$’s and Theorem 1 is proved.
C Proof of inequality (6)

We show that

\[ n^{-1} \text{tr} \left( (T_n(|\lambda|^{-2d_i})^{-1}T_n(f_i - f_0))^2 \right) \geq B \int (f_0/f_i - 1)^2(x)dx \]

when \( d_i > d_0 \) First we prove that

\[ n^{-1} \text{tr} \left( (T_n(|\lambda|^{2d_i}(f_i - f_0))^2 \right) \geq \frac{1}{2} \int |\lambda|^{4d_i}(f_0 - f_i)^2(x)dx \]

when \( d_0 > d_i \) uniformly in \( f_i \in \mathcal{F}_n \) when \( n \) is large enough. We have,

\[
\Delta = n^{-1} \text{tr} \left( (T_n(|\lambda|^{2d_i}(f_i - f_0))^2 - n^{-1} \text{tr} \left( (T_n(|\lambda|^{4d_i}(f_i - f_0))^2 \right) \right) \\
= n^{-1} \int_{[-\pi,\pi]^2} g(\lambda_1)(g(\lambda_2) - g(\lambda_1))\Delta_n(\lambda_1 - \lambda_2)\Delta_n(\lambda_2 - \lambda_1)d\lambda_1d\lambda_2
\]

where

\[
g(\lambda) = |\lambda|^{2d_i}(f_i - f_0)(\lambda) = \tilde{f}_i(\lambda) - |\lambda|^{2(d_i - d_0)}\tilde{f}_0(\lambda).
\]

The second term belongs to either \( \tilde{G}(t, M_0, m_0, L_0, \rho) \) or \( \tilde{\cup}_d\mathcal{L}(d, M_0, m_0, L_0) \), hence this term is treated as in Appendix A. The only term that causes problem is the difference \( \tilde{f}_i(\lambda_1) - \tilde{f}_i(\lambda_2) \) since the derivative of \( \tilde{f}_i \) is not uniformly bounded. Note however that

\[
|\tilde{f}_i(\lambda_1) - \tilde{f}_i(\lambda_2)| \leq (\sum_{j=0}^{k} j|\theta_j|)|\lambda_1 - \lambda_2|
\]

in a FEXP(k) model. So that

\[
\begin{align*}
I_1 &= n^{-1} \left| \int_{[-\pi,\pi]^2} g(\lambda_1)(g(\lambda_2) - g(\lambda_1))\Delta_n(\lambda_1 - \lambda_2)\Delta_n(\lambda_2 - \lambda_1)d\lambda_1d\lambda_2 \right| \\
&\leq n^{-1} \left( \sum_{j=0}^{k} j|\theta_j| \right) \int_{[-\pi,\pi]^2} |g(\lambda_1)|L_n(\lambda_1 - \Lambda_2)d\lambda_2d\lambda_1 \\
&\leq 2\pi \left( \sum_{j=0}^{k} j|\theta_j| \right) \log nM \\
&\leq 40
\end{align*}
\]
In the set $\mathcal{F}_n$ $k \leq k_0 \log n/n$ where $k_0$ is chosen as small as need be. Using $\sum_{j=0}^{k} |\theta_j| \leq k_0 A \log n/n$, we obtain that

$$I_1 \leq \delta,$$

where $\delta$ is chosen as small as we want, depending on $k_0$. Now, we also have that using the same representation of the traces:

$$I_2 = \frac{1}{n} \left[ \text{tr} \left( T_n(|\lambda|^{2d_i})^{-1} T_n(f_i - f_0) \right)^2 \right] - \text{tr} \left( T_n(|\lambda|^{2d_i}(f_i - f_0))^2 \right) \leq \frac{2M^2}{n} \int_{|\lambda| \leq 1} |\lambda_2 \lambda_4|^{-2(d_1 \vee d_0)} \left| |\lambda_1|^{2d_i} - |\lambda_2|^{2d_i} \right|^2 L_n(\lambda_1 - \Lambda_2) \cdots L_n(\lambda_4 - \Lambda_1) d\lambda_1 \cdots d\lambda_4.$$

Using the same calculations as Dahlhaus (1989, p1760-1761), since

$$\left| |\lambda_1|^{2d_i} - |\lambda_2|^{2d_i} \right| \leq K \frac{|\lambda_1 - \lambda_2|^{1-3\delta}}{|\lambda_2|^{1-2\delta - \delta}}, \forall \delta > 0$$

we have that

$$I_2 \leq Kn^{6\delta-1} \log n^3 \to 0, \text{ when } \delta < 1/6.$$

Finally consider

$$I_3 = n^{-1} \left[ \text{tr} \left( T_n(|\lambda|^{2d_i})^{-1} T_n(f_i - f_0) \right)^2 \right] - \text{tr} \left( T_n(|\lambda|^{2d_i}(f_i - f_0))^2 \right) \leq Cn^{\delta-1},$$

for any $n > 0$, using Dahlhaus’ proof of Theorem 5.1. (1989, p 1762). Putting all these results together we finally obtain that

$$n^{-1} \text{tr} \left( T_n(|\lambda|^{2d_i})^{-1} T_n(f_i - f_0) \right)^2 \geq \int |\lambda|^{4d_i}(f_0 - f_i)^2(\lambda)d\lambda - \delta + o(1)$$

where $\delta$ is as small as need be and comes from $I_1$. Since

$$\int |\lambda|^{4d_i}(f_0 - f_i)^2(\lambda)d\lambda > \epsilon |\log \epsilon|^{-2}$$

the result follows.
Lemma 3

Let \( f_j, j \in \{1, 2\} \) be such that \( f_j(\lambda) = |\lambda|^{-2i\delta_j} \tilde{f}_j(\lambda) \), where \( \delta_j < 1/2 \) and \( \tilde{f}_j \in S(L, \beta) \), for some constant \( L > 0 \) and consider \( b \) a bounded function on \([-\pi, \pi]\). Assume that \( h_n(f_1, f_2) < \epsilon \) where \( \epsilon > 0 \). Then \( \forall \delta > 0 \), there exists \( \epsilon_0 > 0 \) such that if \( \epsilon < \epsilon_0 \), there exists \( C > 0 \) such that

\[
\frac{1}{n} \text{tr} \left( T_n(f_1)^{-1}T_n(f_1b)T_n(f_2)^{-1}T_n(f_1b) \right) \leq C (\log n)^3 |b|_2^2 + C |b|_\infty^2 n \delta^{-1} |b|_\infty^2, \tag{25}
\]

\[
\frac{1}{n} \text{tr} \left( T_n(f_1^{-1})T_n(f_1 - f_2)T_n(f_2^{-1})T_n(f_1 - f_2) \right) \leq Ch_n(f_1, f_2). \tag{26}
\]

Let \( g_j = (1 - \cos \lambda)^{d_j} \) and \( f_j = g_j^{-1} \tilde{f}_j \), where \( \tilde{f}_1 \in S(L, \beta) \cap \mathcal{L} \) and \( \tilde{f}_2 \in S(L, \beta) \), written in the form \( \log \tilde{f}_2(\lambda) = \sum_{l=0}^{K} \theta_l \cos l\lambda \),

\[
\left| \frac{1}{n} \text{tr} \left( T_n(g_1(f_1 - f_2))T_n(g_2(f_1 - f_2)) \right) - \text{tr} \left( T_n(g_1g_2(f_1 - f_2)^2) \right) \right|
\leq C n^{-1+\delta} + n^{-1} \log n \sum_{l=0}^{K} |\theta_l| \left( \int_{[-\pi, \pi]} g_1g_2(f_1 - f_2)^2(\lambda)d\lambda \right)^{1/2}, \tag{27}
\]

for any \( \delta > 0 \).

**Proof.** Throughout the proof \( C \) denotes a generic constant. We first prove (25). To do so, we obtain an upper bound on another quantity, namely

\[
\gamma(b) = \frac{1}{n} \text{tr} \left( T_n(f_1^{-1})T_n(f_1b)T_n(f_2^{-1})T_n(f_1b) \right). \tag{28}
\]

Let \( \Delta_n(\lambda) = \sum_{j=1}^{n} \exp(-i\lambda j) \) and \( L_n \) be the \( 2\pi \) periodic function defined by \( L_n(\lambda) = n \) if \( |\lambda| \leq 1/n \) and \( L_n(\lambda) = |\lambda|^{-1} \) if \( 1/n \leq |\lambda| \leq \pi \). Then \( |\Delta_n(\lambda)| \leq CL_n(\lambda) \) and we can express
traces of products of Toeplitz matrices in the following way:

\[
\gamma(b) = C \int_{[-\pi,\pi]^4} \frac{f_1(\lambda_1)b(\lambda_1) f_1(\lambda_3)b(\lambda_3)}{f_0(\lambda_2)f_0(\lambda_4)} \times \\
\Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_3) \Delta_n(\lambda_3 - \lambda_4) \Delta_n(\lambda_4 - \lambda_1) d\lambda_1 \ldots d\lambda_4 \\
= \frac{C}{n} \int_{[-\pi,\pi]^4} \frac{f_1(\lambda_1)f_1(\lambda_3)}{f_0(\lambda_2)f_0(\lambda_4)} (b(\lambda_1)^2 + b(\lambda_1)b(\lambda_3) - b(\lambda_1)^2) \\
\times \Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_3) \Delta_n(\lambda_3 - \lambda_4) \Delta_n(\lambda_4 - \lambda_1) d\lambda_1 \ldots d\lambda_4 \\
= \frac{C}{n} \text{tr} (T_n(f_1b^2)T_n(f_1^{-1})T_n(f_1T_n(f_2^{-1}))) \\
+ \frac{C}{n} \int_{[-\pi,\pi]^4} \frac{f_1(\lambda_1)f_1(\lambda_3)b(\lambda_1)}{f_1(\lambda_2)f_2(\lambda_4)} (b(\lambda_3) - b(\lambda_1)) \\
\times \Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_3) \Delta_n(\lambda_3 - \lambda_4) \Delta_n(\lambda_4 - \lambda_1) d\lambda_1 \ldots d\lambda_4.
\]

On the set \(b(\lambda_1) > b(\lambda_3), 0 < b(\lambda_1) - b(\lambda_3) < b(\lambda_1)\) and on the set \(b(\lambda_3) > b(\lambda_1), 0 < b(\lambda_3) - b(\lambda_1) < b(\lambda_3)\), therefore the second term of the rhs of the above inequality is bounded by (in absolute value)

\[
\gamma_2(b) \leq \frac{2}{n} \int_{[-\pi,\pi]^4} \frac{f_1(\lambda_1)f_1(\lambda_3)b(\lambda_1)^2}{f_1(\lambda_2)f_2(\lambda_4)} L_n(\lambda_1 - \lambda_2)L_n(\lambda_2 - \lambda_3) \\
\times L_n(\lambda_3 - \lambda_4)L_n(\lambda_4 - \lambda_1) d\lambda_1 \ldots d\lambda_4 \\
\leq \frac{C}{n} \int_{[-\pi,\pi]^4} b(\lambda_1)^2 \frac{\lambda_1^{2d_1} \lambda_3^{2d_1}}{\lambda_2^{2d_1} \lambda_4^{2d_1}} L_n(\lambda_1 - \lambda_2)L_n(\lambda_2 - \lambda_3) \\
\times L_n(\lambda_3 - \lambda_4)L_n(\lambda_4 - \lambda_1) d\lambda_1 \ldots d\lambda_4.
\]

Note that

\[
\int_{[-\pi,\pi]} L_n(\lambda_1 - \lambda_2)L_n(\lambda_2 - \lambda_3) d\lambda_2 \leq C \log n L_n(\lambda_1 - \lambda_3),
\]
\[
\gamma_2(b) \leq C (\log n)^3 \int_{[-\pi, \pi]^4} b(\lambda)^2 d\lambda \\
+ C \int_{[-\pi, \pi]^4} b(\lambda_1)^2 |\lambda_1|^{-2(d_1 - d_2)} \left( \frac{|\lambda_3|^{-2d_1}}{|\lambda_2|^{-2d_1}} - 1 \right) \left( \frac{|\lambda_1|^{-2d_2}}{|\lambda_4|^{-2d_2}} - 1 \right) \\
\times L_n(\lambda_1 - \lambda_2) L_n(\lambda_2 - \lambda_3) L_n(\lambda_3 - \lambda_4) L_n(\lambda_4 - \lambda_1) d\lambda_1 \cdots d\lambda_4 \\
+ 2C \int_{[-\pi, \pi]^4} b(\lambda_1)^2 |\lambda_1|^{-2(d_1 - d_2)} \left( \frac{|\lambda_3|^{-2d_1}}{|\lambda_2|^{-2d_1}} + \frac{|\lambda_1|^{-2d_2}}{|\lambda_4|^{-2d_2}} - 2 \right) \\
\times L_n(\lambda_1 - \lambda_2) L_n(\lambda_2 - \lambda_3) L_n(\lambda_3 - \lambda_4) L_n(\lambda_4 - \lambda_1) d\lambda_1 \cdots d\lambda_4.
\]

Since
\[
\left| \frac{|\lambda_1|^{-2d_j}}{|\lambda_2|^{-2d_j}} - 1 \right| \leq C \frac{|\lambda_1 - \lambda_2|^{1-\delta}}{|\lambda_1|^{1-\delta}}, \text{ for } j = \{1, 2\}, \quad (29)
\]

using Dahlhaus (1989)’s (1989) calculations as in his proof of Lemma 5.2, we obtain that, if \(d_1 - d_2 < \delta/4\),
\[
\int_{[-\pi, \pi]^4} b(\lambda_1)^2 |\lambda_1|^{-2(d_1 - d_2)} \left( \frac{|\lambda_3|^{-2d_1}}{|\lambda_2|^{-2d_1}} - 1 \right) \left( \frac{|\lambda_1|^{-2d_2}}{|\lambda_4|^{-2d_2}} - 1 \right) \\
\times L_n(\lambda_1 - \lambda_2) L_n(\lambda_2 - \lambda_3) L_n(\lambda_3 - \lambda_4) L_n(\lambda_4 - \lambda_1) d\lambda_1 \cdots d\lambda_4 \\
\leq |b|_\infty^2 \int_{[-\pi, \pi]^4} |\lambda_1|^{-1+\delta/2} |\lambda_4|^{-1+\delta} L_n(\lambda_1 - \lambda_2) L_n(\lambda_2 - \lambda_3)^\delta L_n(\lambda_3 - \lambda_4) L_n(\lambda_4 - \lambda_1)^\delta d\lambda_1 \cdots d\lambda_4 \\
\leq C n^{2\delta} |b|_\infty^2 (\log n)^2,
\]
as soon as \(|d_1 - d_2| < \delta/2\). By considering \(h_n(f, f_0) < \epsilon\) with \(\epsilon > 0\) small enough, we can impose that \(|d_1 - d_2| < \delta/2\), and we finally obtain that
\[
\gamma(b) \leq C |b|_2^3 (\log n)^3 + C |b|_\infty^2 n^{2\delta-1} (\log n)^2 |b|_\infty^2.
\]
and (25) is proved. We now prove (26). Since $f_j \geq m|\lambda|^{-2d_j} = g_j$ where $m = e^{-L}$, $T_n^{-1}(f_j) < T_n^{-1}(g_j)$, i.e. $T_n^{-1}(g_j) - T_n^{-1}(f_j)$ is positive semi definite, and

$$h_n(f_1, f_2) = \frac{1}{2n} \text{tr} \left( T_n(f_1 - f_2) T_n^{-1}(f_2) T_n(f_1 - f_2) T_n^{-1}(f_1) \right)$$

$$\geq \frac{1}{2n} \text{tr} \left( T_n(f_1 - f_2) T_n^{-1}(f_2) T_n(f_1 - f_2) T_n^{-1}(g_1) \right)$$

$$\geq \frac{1}{2n} \text{tr} \left( T_n(f_1 - f_2) T_n^{-1}(f_2) T_n(f_1 - f_2) T_n^{-1/2}(g_1) R_1 T_n^{-1/2}(g_1) \right)$$

$$+ \frac{1}{2n} \text{tr} \left( T_n(f_1 - f_2) T_n^{-1}(g_2) T_n(f_1 - f_2) T_n \left( g_1^{-1} \right) \right)$$

$$= \frac{1}{2n(16\pi^2)} \text{tr} \left( T_n(f_1 - f_2) T_n(g_2^{-1}) T_n(f_1 - f_2) T_n \left( g_1^{-1} \right) \right)$$

$$+ \frac{1}{2n} \text{tr} \left( T_n(f_1 - f_2) T_n^{-1/2}(g_2) R_2 T_n^{-1/2}(g_2) T_n(f_1 - f_2) T_n \left( g_1^{-1} \right) \right)$$

$$\geq \frac{1}{2n(4\pi^2)} \text{tr} \left( T_n(f_1 - f_2) T_n^{-1/2}(g_2) R_2 T_n^{-1/2}(g_2) T_n(f_1 - f_2) T_n \left( g_1^{-1} \right) \right)$$

(30)

where $R_j = \text{id} - T_n^{1/2}(g_j) T_n(g_j^{-1}/(4\pi^2)) T_n^{1/2}(g_j)$. We first bound the first term of the rhs of (30). Let $\delta > 0$ and $\epsilon < \epsilon_0$ such that, $|d - d_0| \leq \delta$ (Corollary 1 implies that there exists such a $\epsilon_0$). Then using Lemmas 5.2 and 5.3 of Dahlhaus (1989)

$$\left| \text{tr} \left( T_n(f_1 - f_2) T_n^{-1}(f_2) T_n(f_1 - f_2) T_n^{-1/2}(g_1) R_1 T_n^{-1/2}(g_1) \right) \right|$$

$$\leq 2 |R_1| |T_n^{1/2}(g_1) T_n(f_1 - f_2) T_n^{-1/2}(f_2)| |T_n(|f_1 - f_2|)^{1/2} T_n^{-1/2}(f_2)| |T_n(|f_1 - f_2|)^{1/2} T_n^{-1/2}(g_1)|$$

$$\leq C n^{3h} |T_n^{1/2}(g_1) T_n(f_1 - f_2) T_n^{-1/2}(f_2)|.$$

Since $g_1 \leq Cf_1$,

$$|T_n^{1/2}(g_1) T_n(f_1 - f_2) T_n^{-1/2}(f_2)|^2 = \text{tr}(T_n^{-1}(g_1) T_n(f_1 - f_2) T_n^{-1}(f_2) T_n(f_1 - f_2))$$

$$\leq C \text{tr}(T_n^{-1}(f_1) T_n(f_1 - f_2) T_n^{-1}(f_2) T_n(f_1 - f_2))$$

$$= C n h_n(f_1, f_2),$$

and

$$\frac{1}{n} \left| \text{tr} \left( T_n(f_1 - f_2) T_n^{-1}(f_2) T_n(f_1 - f_2) T_n^{-1/2}(g_1) R_1 T_n^{-1/2}(g_1) \right) \right| \leq C n^{25 - 1/2} h_n(f_1, f_2).$$
We now bound the second term of the rhs of (30).

\[
\begin{align*}
&= \left| \frac{1}{n} \operatorname{tr} \left( T_n(f_1-f_2)T_n^{-1/2}(g_2)R_2T_n^{-1/2}(g_2)T_n(f_1-f_2)T_n(g_1^{-1}) \right) \right| \\
&\leq \frac{1}{n} |R_2| T_n^{-1/2}(g_2)T_n(f_1-f_2)T_n(g_1)^{-1/2} T_n(g_1)^{1/2} T_n(f_1-f_2) T_n^{-1/2}(f_2) \\
&\leq \frac{C n^\delta \sqrt{n h_n(f_2, f_1)}}{n} \left\| T_n(g_1)^{1/2} T_n(g_1^{-1}) T_n(f_1-f_2) T_n^{-1/2}(f_2) \right\| \\
&\leq \frac{C n^{\delta+1/2} \sqrt{n h_n(f_2, f_1)}}{n} \left\| T_n(g_1)^{1/2} T_n(g_1^{-1}) \right\|^2 \\
&\leq C n^{3\delta-1/2} h_n(f_1, f_2),
\end{align*}
\]

since \(\left\| T_n(f_1)^{1/2} T_n(f_1^{-1}) \right\| \leq \| \text{id} \| + \left\| T_n(f_1)^{1/2} T_n(f_1^{-1}) T_n(f_1)^{1/2} - \text{id} \right\| \leq C n^\delta\).

Therefore,

\[
\frac{C}{n} \operatorname{tr} \left( T_n(f_1-f_2)T_n(g_2^{-1})T_n(f_1-f_2)T_n(g_1^{-1}) \right) \leq C h_n(f_1, f_2)(1 + n^{-1/2+3\delta}),
\]

and using the fact that \(g_j^{-1} > C f_j^{-1}\), for \(j = 1, 2\) we prove (26).

The proof of (27) is similar:

\[
A = \operatorname{tr} \left( T_n(g_1(f_1-f_2))T_n(g_2(f_1-f_2)) \right) - \operatorname{tr} \left( T_n(g_1 g_2(f_1-f_2)^2) \right)
\]

\[
= C \int_{[-\pi,\pi]^2} g_1(f_1-f_2)(\lambda_1) [g_2(f_1-f_2)(\lambda_2) - g_1(f_1-f_2)(\lambda_1)] \Delta_n(\lambda_1-\lambda_2) \ldots \Delta_n(\lambda_4-\lambda_1) d\lambda
\]

\[
- C \int_{[-\pi,\pi]^2} g_1 g_2(f_1-f_2)(\lambda_1) [f_1(\lambda_2) - f_1(\lambda_1)] \Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_1) d\lambda
\]

The first 2 terms of the right hand side are of order \(O(n^{2\delta} \log n)\). We now study the last term, here the problem is due to the fact that \(\hat{f}_2\) does not necessarily belong to \(\mathcal{L}\). We have
\[
\int_{[-\pi,\pi]^2} g_1 g_2 (f_1 - f_2)(\lambda_1) [f_2(\lambda_2) - f_2(\lambda_1)] \Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_1) d\lambda
\]

\[
= \int_{[-\pi,\pi]^2} g_1 g_2 (f_1 - f_2)(\lambda_1) \tilde{f}_2(\lambda_2) \Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_1) d\lambda
\]

\[
+ \int_{[-\pi,\pi]^2} g_1 (f_1 - f_2)(\lambda_1) \tilde{f}_2(\lambda_2) - \tilde{f}_2(\lambda_1) \Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_1) d\lambda.
\]

The first term of the above inequality is of order \(O(n^{25} \log n)\) since \(g_2\) belongs to \(L\). Since

\[
\tilde{f}(\lambda) = \exp\{\sum_{l=0}^{K_n} \theta_l \cos l \lambda\},
\]

\[
I = \int_{[-\pi,\pi]^2} g_1 (f_1 - f_2)(\lambda_1) [\tilde{f}_2(\lambda_2) - \tilde{f}_2(\lambda_1)] \Delta_n(\lambda_1 - \lambda_2) \Delta_n(\lambda_2 - \lambda_1) d\lambda
\]

\[
\leq C \int_{[-\pi,\pi]^2} g_1 |f_1 - f_2|(\lambda_1) \left| \sum_{j=0}^{K_n} \theta_j (\cos (j \lambda_2) - \cos (j \lambda_1)) \right| L_n(\lambda_1 - \lambda_2) L_n(\lambda_2 - \lambda_1) d\lambda
\]

\[
\leq C \log n \left( \sum_{l=0}^{K_n} |\theta_l| l \right) \int_{[-\pi,\pi]} g_1 |f_1 - f_2|(\lambda) d\lambda
\]

\[
\leq C \log n \sum_{l=0}^{K_n} |\theta_l| l \left( \int_{[-\pi,\pi]} g_1 g_2 (f_1 - f_2)^2(\lambda) d\lambda \right)^{1/2},
\]

where the latter inequality holds because \(\int g_1 / g_2(\lambda) d\lambda\) is bounded and via an application of Hölder inequality.

References


**Acknowledgements**

Part of this work was done while the second Author was visiting the Université Paris Dauphine, CEREMADE. He thanks for warm hospitality and financial support.
Judith Rousseau

Ceremade

Université Paris Dauphine

Place du Marchal De Lattre, de Tassigny, 75016

Paris, France

e-mail: rousseau@ceremade.dauphine.fr

Brunero Liseo

Dip. studi geoeconom., linguist., statist.

Stor. per l'analisi regionale

Università di Roma “La Sapienza”

Via del Castro Laurenziano, 9

I-00161 Roma, Italia

e-mail: brunero.liseo@uniroma1.it