Marginal density estimation for linear processes with seasonal long memory

Mohamedou OULD HAYE[1] and Anne PHILIPPE[2]

[1]School of Mathematics and Statistics, Carleton University
1125 Colonel By Drive
Ottawa, Ontario K1S 5B6, Canada
[1] Laboratoire de Mathmatiques Jean Leray UMR CNRS 6629
Universit de Nantes
2 rue de la Houssinire - BP 92208 44322 Nantes Cedex 3, France

Abstract

Some convergence results on the kernel density estimator are proven for a class of linear processes with seasonal effects. In particular we extend the results of Ho and Hsing (1996a) and Mielniczuk (1997); Hall and Hart (1990) to the stationary processes for which the singularities of the spectral density are not limited to the origin. We show that the convergence rates and the limit distribution may be different in this context.

Keyword : Confidence band ; empirical process ; limit theorem ; mean integrated squared error.

1 Introduction

Hosking (1981) introduced long memory processes with quasi periodic behaviour. This fact corresponds, for stationary processes, to spectral densities which exhibit singularities at non zero frequencies. Many authors have contributed to the construction of fractional models with singularities/poles outside the origin, see for instance, Gray et al. (1994, 1989); Hassler (1994); Viano et al. (1995); Leipus and Viano (2000); Bisognin and Lopes (2009).

We can distinguish between two types of long memory: one regular and the other seasonal according to whether the spectral density has a pole at the origin or outside the origin. From a statistical point of view, the estimators of the long memory parameter have been adapted to yield some estimates if seasonal effects are assumed. In a parametric context, the $\sqrt{n}$-consistency of the maximum likelihood estimate or the Whittle estimate has been proved (see Hosoya (1997); Giraitis et al. (2001) when the pole is unknown). Semi parametric estimates can be more or less easily adapted to the *seasonal* case (see Hidalgo and Soulier (2004); Arteche and Robinson (2000, 1999); Hsu and Tsai (2009); Reisen et al. (2006); Whitcher (2004)).

When we consider empirical process related statistics the situation is more delicate. The normalisation and the limit distribution can be different according to whether the memory is regular or seasonal. An important literature is devoted to the convergence of the empirical process, see for instance Ho and Hsing (1996b); Giraitis and Surgailis (1999) in regular case and Ould Haye (2002) Ould Haye and Philippe (2003) in seasonal case.

In this paper we give some convergence results on the kernel estimator of the marginal density f. Let $(X_1, \cdots, X_n)$ be an observed sample from f, the kernel estimator of f is defined by

$$\tilde{f}_n(x) = \frac{1}{nm_n} \sum_{j=1}^{n} K\big(\frac{x - X_j}{m_n}\big). \tag{1.1}$$

where m_n is the bandwidth and K is a kernel function.

Consider the following infinite moving average process,

$$X_n = \sum_{j=-\infty}^{n} b(n-j)\xi_j, \qquad n \geq 1 \tag{1.2}$$

where

- the sequence $(b(k))_k$ has the form

$$b(n) = n^{-(\alpha+1)/2} \sum_{j \in J} a_j \big(\cos n\lambda_j + o(1)\big), \tag{1.3}$$

  where $\alpha \in (0,1)$ and $\lambda_j \neq 0$ for all $j \in J$ a finite subset of $\mathbb{N}$.

- $(\xi_n)_n$ is a sequence of independent and identically distributed random variables with zero mean and finite variance $\mathbb{E}\xi_0^2 = \sigma^2 < \infty$.

From Giraitis and Leipus (1995), the covariance function r of (X_t) defined in (1.2) has the form

$$r(n) = n^{-\alpha} \sum_{j \in J} a_j \big(\cos n\lambda_j + o(1)\big). \tag{1.4}$$

Note that the condition on the coefficient α ensures that $\sum |r(n)| = \infty$, thus the process has a long-memory. But $\sum r(n)$ may be finite, and in any case $|\sum r(n)| = o\big(\sum r(n)^2\big)$. This fact characterises seasonal long memory and the asymptotic behavior of many statistics (see below for the empirical process) can be drastically different when $\alpha < 1/2$. This is mainly due to the fact that X_t^2 will also have a long memory. We focus on this case.

A large class of linear processes satisfying these conditions is obtained by filtering a white noise (ξ_i) as follows:

$$X_t = G(B)\xi_t \text{ with } G(z) = g(z) \prod_{j=-m}^{m} \big(1 - e^{i\lambda_j} z\big)^{(\alpha_j - 1)/2}, \quad m \geq 1, \tag{1.5}$$

where B is the backshift operator and where g is an analytic function on $\{|z| < 1\}$, continuous on $\{|z| \leq 1\}$ and $g(z) \neq 0$ if $|z| = 1$. Indeed taking

$$0 < \alpha_j < 1, \quad \alpha_j = \alpha_{-j}, \quad \lambda_{-j} = -\lambda_j, \quad j = 0, \ldots, m, \text{ and}$$

$$0 = \lambda_0 < \lambda_1 < \ldots < \lambda_m < \pi,$$

the condition (1.3) is then satisfied with

$$\alpha = \min\{\alpha_j,\ j = 0, \ldots, m\}, \quad J = \{j \geq 0 \ :\ \alpha_j = \alpha\}.$$

We consider the empirical process associated with the process $(X_n)_{n \geq 1}$ defined by

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{\{X_j \leq x\}}.$$

Ould Haye and Philippe (2003) proved the following results for the linear process (X_n) defined in (1.5). Assume that $\mathbb{E}\xi_0^4 < \infty$, the cumulative distribution function of ξ_0 is 5 times differentiable with continuous bounded and integrable derivatives on $\mathbb{R}$. Denote

$$d_n = n^{1-\alpha}, \quad \text{and} \quad D = \frac{\sqrt{(2-2\alpha)(1-2\alpha)}}{4\Gamma(\alpha)\cos(\alpha\pi/2)}.$$

Assume that $\alpha < 1/2$.
Then, as n tends to infinity,

$$d_n^{-1}[nt]\big(F_{[nt]}(x) - F(x)\big) \Longrightarrow \frac{F''(x)}{2} R(t), \tag{1.6}$$

where R is a linear combination of independent Rosenblatt processes with the same parameter α

$$R(t) = R_{\alpha,\Lambda}(t) = D^{-1} \sum_{j\in J} c_j \Big(R_j^{(1)}(t) + R_j^{(2)}(t)\Big), \tag{1.7}$$

where $\Lambda = \{\lambda_j, \quad j \in J\}$, and where

- $c_0 = h_0/2$, $c_j = h_j$ if $j \neq 0$ and

$$h_j = g(e^{i\lambda_j}) \prod_{\ell\neq j} \big(1 - e^{i(\lambda_\ell - \lambda_j)}\big)^{(\alpha-1)/2},$$

- $R_j^{(i)}(t)$, $i = 1, 2$ and $j \in J$ are Rosenblatt processes with parameter $1 - \alpha$, independent except for $j = 0$, $R_0^{(1)}(t) = R_0^{(2)}(t)$.

The paper is organized as follows. In Section 2, we establish a limit theorem for the kernel estimate. This extends one of Ho and Hsing (1996a)'s results, in particular we show the contribution and the effect of the singularities of the spectral density outside the origin to the convergence rate and the limit distribution. Then we apply our limit theorem to construct confidence bands for the density function.

Similarly to Hall and Hart (1990); Mielniczuk (1997), we provide in Section 3, the asymptotic behavior of the mean integrated squared error, and we show that the equivalence can be modified when the singularities of the spectral density are not limited to the origin.

2 Asymptotic distribution of the kernel estimator

Hereafter, we assume that the kernel K is a continuous function with compact support and $\int K(x)dx = 1$. Concerning the bandwidth m_n, we assume that $m_n \to 0$ and $nm_n \to \infty$, as n tends to infinity.
The equality

$$\tilde{f}_n(x) - \mathbb{E}\tilde{f}_n(x) = \frac{1}{m_n} \int_{\mathbb{R}} K\big(\frac{x-u}{m_n}\big) d\big(F_n(u) - F(u)\big) \tag{2.8}$$

clearly shows the relationship between the estimate $\tilde{f}_n(x)$ and the empirical process $F_n(x)$. The process $\tilde{f}_n(x)$ is sometimes called the empirical density process.

For every integer $n \geq 1$, We define the following statistics

$$Y_{n,1} = \sum_{k=1}^{n} X_k, \qquad Y_{n,2} = \sum_{k=1}^{n} \sum_{s<r} b_r b_s \xi_{k-s} \xi_{k-r}, \tag{2.9}$$

and

$$S_{n,2}(x) = n\big(F_n(x) - F(x)\big) + F'(x) Y_{n,1} - \frac{1}{2} F''(x) Y_{n,2}. \tag{2.10}$$

Remark 1 For linear processes defined in (1.5), the following equivalences as n tends to infinity, have been proved by Ould Haye and Philippe (2003)

$$\mathrm{Var}(Y_{n,2}) \sim \frac{1}{4} \mathrm{Var}\Big(\sum_{j=1}^{n} (X_j^2 - \mathbb{E}(X_1^2))\Big) \sim Cn^{2-2\alpha}. \tag{2.11}$$

and

$$\mathrm{Var}(Y_{n,1}) = \mathrm{Var}\Big(\sum_{j=1}^{n} X_j\Big) \sim C n^{2-\alpha_0}. \tag{2.12}$$

Therefore (2.12) and (2.11) imply that the convergence rate obtained in Proposition 2.1 is smaller than the convergence rate of $\bar{X}_n$.

Let us define the class of Parzen kernels of order s.

Definition 2.1 *A kernel function K is said to be a Parzen kernel of order $s \geq 2$ if it satisfies the following conditions*

1. $\int_{\mathbb{R}} K(u)du = 1$,
2. *for every* $1 \leq j \leq s-1$, $\int_{\mathbb{R}} u^j K(u)du = 0$,
3. $\int_{\mathbb{R}} |u^s||K(u)|du < \infty$.

Bretagnolle and Huber (1979) proved the existence of such kernels, for which, an explicit construction can be found in Gasser and Müller (1979).

Proposition 2.1 *Consider a process (X_n) defined in (1.5). Assume that $\mathbb{E}\xi_0^4 < \infty$, the cumulative distribution function of ξ_0 is 5 times differentiable with continuous bounded and integrable derivatives on $\mathbb{R}$. Assume that $\alpha < (1 \wedge \alpha_0)/2$. Let K be a Parzen kernel of order 4 having bounded total variation. Assume that the bandwidth has the form*

$$m_n = n^{-\delta}, \quad \textit{where } \frac{\alpha}{4} < \delta < \frac{\alpha}{2}.$$

Then, as n tends to infinity

$$n^{\alpha} \sup_{x \in \mathbb{R}} |\tilde{f}_n(x) - f(x)| \xrightarrow{d} \sup_{x \in \mathbb{R}} \Big|\frac{f''(x)}{2}\Big| |R_{\alpha,\Lambda}|. \tag{2.13}$$

where $R_{\alpha,\Lambda} = R_{\alpha,\Lambda}(1)$. Moreover,

$$n^{\alpha}(\tilde{f}_n(x) - f(x)) \overset{C_b(\mathbb{R})}{\Longrightarrow} -\frac{f''(x)}{2} R_{\alpha,\Lambda}, \tag{2.14}$$

where $\overset{C_b(\mathbb{R})}{\Longrightarrow}$ denotes the convergence in $C_b(\mathbb{R})$, the space of continuous bounded functions.

Proof:
The difference between $\tilde{f}_n$ and f can be expressed as

$$\begin{aligned}\tilde{f}_n(x) - f(x) &= \tilde{f}_n(x) - \mathbb{E}\tilde{f}_n(x) + \mathbb{E}\tilde{f}_n(x) - f(x)\\ &= \frac{1}{m_n}\int K(u)d\big(F_n(x - m_n u) - F(x - m_n u)\big) + \int \big(f(x - m_n u) - f(x)\big)K(u)du.\end{aligned}$$

We first replace $F_n - F$ by its expression in (2.10). Then we apply the integration by parts formula on the first integral. For the second, we apply the Taylor-Lagrange formula. There exists a real number u^* such that $|u^* - x| < |m_n u|$ and

$$\begin{aligned}\tilde{f}_n(x) - f(x) =& \frac{-1}{n m_n}\int S_{n,2}(x - m_n u)dK(u) + \frac{Y_{n,1}}{n}\int f'(x - m_n u)K(u)du\\ &- \frac{Y_{n,2}}{n} f''(x)\int K(u)du + \frac{Y_{n,2}}{n} m_n \int f^{(3)}(u^*)uK(u)+\\ &+ \int \Big(-m_n u f'(x) + \frac{m_n^2 u^2}{2} f''(x) - \frac{m_n^3 u^3}{6} f^{(3)}(x) + \frac{m_n^4 u^4}{24} f^{(4)}(u^*)\Big)K(u)du\\ =:& a_n(x) + b_n(x) + c_n(x) + d_n(x) + e_n(x).\end{aligned}$$

Now, a proof similar to that of Theorem 2.2 in Ho and Hsing (1996a) allows us to write for $2\delta < \alpha$

$$n^{\alpha+\delta-1}\sup_{x\in\mathbb{R}}|S_{n,2}(x)| \xrightarrow{a.s.} 0, \qquad \text{as } n \to \infty. \tag{2.15}$$

And, thus we have

$$n^{\alpha}\sup_{x\in\mathbb{R}}|a_n(x)| \xrightarrow{P} 0, \qquad \text{as } n \to \infty \tag{2.16}$$

where $\xrightarrow{P}$ denotes the convergence in probability.
For the sequences $b_n(x)$, $d_n(x)$, $e_n(x)$, we get the same convergence in probability as in (2.16) by bounding the variances. To obtain the bounds, we start from the variances of $Y_{n,1}$ and $Y_{n,2}$ defined in (2.12) and (2.11), and we use the fact that K is a Parzen kernel and that f is 4 times differentiable and bounded derivatives. We get, as n tends to infinity,

$$\begin{aligned}\mathrm{Var}(n^{\alpha}\sup_{x\in\mathbb{R}}|b_n(x)|) &\leq n^{2\alpha-2}\,\mathrm{Var}\Big(Y_{n,1}\sup_{x\in\mathbb{R}}|f'(x)|\int|K(u)|du\Big)\\ &= Cn^{2\alpha-2}n^{2-\alpha_0} = Cn^{2\alpha-\alpha_0} \longrightarrow 0,\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(n^{\alpha}\sup_{x\in\mathbb{R}}|d_n(x)|) &\leq n^{2\alpha-2}\,\mathrm{Var}\Big(Y_{n,2}m_n\sup_{x\in\mathbb{R}}|f^{(3)}(x)|\int|uK(u)|du\Big)\\ &= Cn^{2\alpha-2}n^{2-2\alpha}n^{-\delta} \longrightarrow 0,\end{aligned}$$

$$n^{\alpha}\sup_{x\in\mathbb{R}}|e_n(x)| \leq \sup_{x\in\mathbb{R}}|f^{(4)}(x)|\frac{n^{\alpha-4\delta}}{24}\int u^4|K(u)|du = O(n^{\alpha-4\delta}) \longrightarrow 0,$$

These four convergences in probability imply that both sequences

$$n^{\alpha}\sup_{x\in\mathbb{R}}|\tilde{f}_n(x) - f(x)| \quad \text{and} \quad n^{\alpha}\sup_{x\in\mathbb{R}}|f''(x)||\frac{Y_{n,2}}{n}| = n^{\alpha}\sup_{x\in\mathbb{R}}|c_n(x)|$$

have the same limit as n tends to infinity. According to Lemma 2.1 in Ould Haye and Philippe (2003), this common limit is equal to

$$\sup_{x\in\mathbb{R}}\Big|\frac{f''(x)}{2}\Big||R_{\alpha,\Lambda}|.$$

Hence (2.13) is proved. According to (2.11), we notice that the rate $n^{-\alpha}$ given in (2.13) is the convergence rate of $n^{-1}\sum_{j=1}^{n}(X_j^2 - \mathbb{E}(X_1^2))$.
Similarly, as n tends to infinity, the finite-dimensional distributions of

$$n^{\alpha}(\tilde{f}_n(x) - f(x)) \quad \text{and} \quad -n^{\alpha}f''(x)\frac{Y_{n,2}}{n} = n^{\alpha}c_n(x)$$

converge simultaneously to the finite-dimensional distributions of $-(f''(x)/2)R_{\alpha,\Lambda}$. This concludes the proof of (2.14) because (2.13) implies the tightness of $n^{\alpha}(\tilde{f}_n(x) - f(x))$.

Remark 2 We clearly see that the choice of the class of Parzen kernels allows the bias $\mathbb{E}\tilde{f}_n(x) - f(x)$ to become negligible. If K is not a Parzen kernel, the contribution of the bias $e_n(x)$ is not negligible with respect to $b_n(x)$. Therefore, (2.13) is false for a standard kernel unless we replace $\tilde{f}_n(x) - f(x)$ by $\tilde{f}_n(x) - \mathbb{E}\tilde{f}_n(x)$ in (2.13).

Remark 3 The result (2.13) in Proposition 2.1 can be applied to obtain a goodness of fit test on the marginal density.

Remark 4 The result (2.13) in Proposition 2.1 provides confidence bands for f which depend on the derivative f''. In general, f'' is not available, and thus the confidence band cannot be calculated. Then f'' can be replaced by its kernel estimate given by

$$\tilde{f}_n''(x) = \frac{1}{nm_n^3}\sum_{j=1}^{n} K''\big(\frac{x-X_j}{m_n}\big).$$

(note that it is necessary to assume that the kernel function K is twice differentiable.)

Proposition 2.2 *Under the same hypotheses as in Proposition 2.1 and if the kernel function K is twice differentiable and its derivative K'' is continuous, then for each interval $[a,b]$ on which f'' is positive, we have*

$$2n^{\alpha}\sup_{x\in[a,b]}\Big|\frac{\tilde{f}_n(x)-f(x)}{\tilde{f}_n''(x)}\Big| \xrightarrow{d} |R_{\alpha,\Lambda}|. \tag{2.17}$$

In other words, as n tends to infinity, for every $t>0$, we have

$$P\Big\{\tilde{f}_n(x) - \frac{t\tilde{f}_n''(x)}{2n^{\alpha}} \le f(x) \le \tilde{f}_n(x) + \frac{t\tilde{f}_n''(x)}{2n^{\alpha}},\ a\le x\le b\Big\} \to P\big\{|R_{\alpha,\Lambda}| < t\big\}. \tag{2.18}$$

In Proposition 2.3, we give a consistent estimate of the quantiles of process $R_{\alpha,\Lambda}$. Using (2.18), this allows us to obtain asymptotic confidence band for the density $f(x)$ which is valid for every $x\in[a,b]$.

Proof :
Let ϕ be the function defined on $C_b(\mathbb{R})$ by

$$\phi(g) = \sup_{x\in[a,b]}\Big|\frac{g(x)}{f''(x)}\Big|$$

Since ϕ is continuous, (2.14) ensures the following convergence :

$$2n^{\alpha}\sup_{x\in[a,b]}\Big|\frac{\tilde{f}_n(x)-f(x)}{f''(x)}\Big| \xrightarrow{d} |R_{\alpha,\Lambda}|, \qquad \text{as } n\to\infty. \tag{2.19}$$

Now, we prove that the difference

$$Y_n(x) := n^{\alpha}\Big(\frac{\tilde{f}_n(x)-f(x)}{f''(x)} - \frac{\tilde{f}_n(x)-f(x)}{\tilde{f}_n''(x)}\Big)$$

satisfies

$$\sup_{x\in[a,b]}|Y_n(x)| \xrightarrow{P} 0, \qquad \text{as } n\to\infty.$$

This convergence is obtained as follows. We rewrite $Y_n(x)$ as

$$|Y_n(x)| = n^{\alpha}\Big|\frac{\tilde{f}_n(x)-f(x)}{f''(x)}\Big|\Big|\frac{\tilde{f}_n''(x)-f''(x)}{\tilde{f}_n''(x)}\Big|.$$

and by (2.19), it is enough to prove that

$$\sup_{x\in\mathbb{R}}\Big|\frac{\tilde{f}_n''(x)-f''(x)}{\tilde{f}_n''(x)}\Big| \xrightarrow{P} 0, \qquad \text{as } n\to\infty. \tag{2.20}$$

The difference between $\tilde{f}_n''$ and f'' can be written as

$$\begin{aligned}\tilde{f}_n''(x) - f''(x) =& \frac{-1}{nm_n^3}\int S_{n,2}(x-m_nu)dK''(u) + \frac{Y_{n,1}}{n}\int f^{(3)}(x-m_nu)K(u)du - \\ &- \frac{Y_{n,2}}{n}\int f^{(4)}(x-m_nu)K(u)du + \int\big(f''(x-hu)-f''(x)\big)K(u)du.\end{aligned}$$

by replacing f with f'' and $\tilde{f}_n$ with $\tilde{f}''_n$ and following the same lines as the proof of Proposition 2.1. Then, we get

$$\sup_{x\in\mathbb{R}}|\tilde{f}''_n(x) - f''(x)| = O\big(n^{-(2\delta\wedge(1-3\delta))}\big).$$

Since $0 < \delta < 1/4$, we have

$$\sup_{x\in\mathbb{R}}|\tilde{f}''_n(x) - f''(x)| \xrightarrow{P} 0,$$

moreover, the derivative f'' satisfies

$$\inf_{x\in[a,b]}|f''(x)| > 0.$$

Thus, we get (2.20). This concludes the proof.

Proposition 2.3 *Fix $\beta \in (0,1)$. Let $c(\alpha,\Lambda,\beta)$ be the quantile of order β of the process $R_{\alpha,\Lambda}$ defined in* (1.7)*. If (α_n,Λ_n) be consistent (in probability) estimators of (α,Λ). then*

$$c(\alpha_n,\Lambda_n,\beta) \xrightarrow{P} c(\alpha,\Lambda,\beta) \tag{2.21}$$

Remark 5 In the references given in the introduction, the parametric and semi parametric methods provide estimators of (α,Λ) which satisfy the condition required in Proposition 2.3.

Proof : We want to show (2.21) which will be obtained if we show that the application $(\gamma,\theta) \mapsto c(\gamma,\theta,\beta)$ is continuous, as $(\alpha_n,\Lambda_n) \xrightarrow{P} (\alpha,\Lambda)$. To prove this continuity we prove that the mappings g,h below are continuous,

$$((0,1)\times[0,\pi]^{|J|},\ |.|) \xrightarrow{g} (C_b(\mathbb{R}),\|.\|) \xrightarrow{h} ((0,1),|.|),$$

where $\|.\|$ is the uniform metric, and in the following decomposition $F_{\gamma,\theta}$ is the distribution function of $R_{\gamma,\theta}$.

$$(\gamma,\theta) \mapsto [g(\gamma,\theta) = F_{\gamma,\theta}] \mapsto [h(F_{\gamma,\theta}) = c(\gamma,\theta,\beta)].$$

Continuity of g can be proved as follows. Consider a deterministic sequence (γ_n,θ_n) such that $(\gamma_n,\theta_n) \to (\gamma,\theta)$ as $n \to \infty$. Then to prove that $F_{\gamma_n,\theta_n} \to F_{\gamma,\theta}$ uniformly it will be enough to show that $R_{\gamma_n,\theta_n} \Longrightarrow R_{\gamma,\theta}$. To obtain the latter weak convergence it will suffice to show that every sequence of Rosenblatt variables (R_{γ_n}) with parameter γ_n converges weakly to a Rosenblatt variable R_γ with parameter γ, as R_{γ_n,θ_n} is a linear combination of independent Rosenblatt variables R_{γ_n} with the coefficients c_j/D that are continuous functions of γ_n,θ_n. We have from Major (1981)

$$R_{\gamma_n} = \int\int_{\mathbb{R}^2} \frac{e^{i(x+y)}-1}{i(x+y)} W_n(dx,dy)$$

where

$$W_n(dx,dy) = |x|^{(\gamma_n-1)/2}|y|^{(\gamma_n-1)/2}W(dx,dy)$$

with $W(dx,dy)$ being the standard Gaussian random measure, and since

$$|x|^{(\gamma_n-1)/2}|y|^{(\gamma_n-1)/2} \to |x|^{(\gamma-1)/2}|y|^{(\gamma-1)/2}$$

then we have the required convergence.
Now to prove the continuity of h it is enough to note that the quantile function is continuous (with respect to the uniform metric) over the class of monotonic continuous distribution functions, i.e. if $\|F_n - F\| \to 0$ then $h(F_n,\beta)) \to h(F,\beta)$. Of course here we do have $\|F_{\gamma_n,\theta_n} - F_{\gamma,\theta}\| \to 0$, as we just established that $R_{\gamma_n,\theta_m} \Longrightarrow R_{\gamma,\theta}$. □

3 Asymptotic mean integrated squared error (MISE)

The mean integrated squared error (MISE) of the estimate $\tilde{f}_n$ is defined by

$$\int_{\mathbb{R}} \mathbb{E}\big(\tilde{f}_n(x) - f(x)\big)^2 dx.$$

For a wide class of linear processes including the processes with short and *regular* long memories, Hall and Hart (1990) and Mielniczuk (1997) studied the asymptotic behavior of the MISE. In particular, they established the following equivalence, when n tends to infinity,

$$\int_{\mathbb{R}} \mathbb{E}\big(\tilde{f}_n(x) - f(x)\big)^2 dx \sim \int_{\mathbb{R}} \mathbb{E}_0\big(\tilde{f}_n(x) - f(x)\big)^2 dx + \mathrm{Var}(\overline{X}_n) \int_{\mathbb{R}} f'(x)^2 dx \tag{3.22}$$

where $\mathbb{E}_0$ denotes the expectation with respect to the distribution of n independent random variables distributed from the density f. In particular, the equivalence (3.22) shows that the convergence rate of the MISE cannot be faster than the convergence of $\mathrm{Var}(\overline{X}_n)$. In other words, the convergence rate of the kernel density estimates is bounded from above by the convergence rate of the empirical mean. This is the optimal rate.

Hereafter, we assume that the distribution of the innovation (ξ_k) satisfies

[Z1] $E|\xi_1|^m < \infty$

[Z2] for some $\delta > 0$ and $C < \infty$ the characteristic function of ξ_1 satisfies

$$|Ee^{iu\xi_1}| \leq C(1 + |u|)^{-\delta} \tag{3.23}$$

Theorem 3.1 *Let (X_n) be a linear process defined in (1.2) and (1.5) such that $\mathbb{E}\xi_0^4 < \infty$. Assume that $\alpha < (1 \wedge \alpha_0)/3$ and the kernel K is a bounded symmetric density function. Then the MISE satisfies, as n tends to infinity,*

$$MISE(\tilde{f}_n) \sim \int_{\mathbb{R}} \mathbb{E}_0\big(\tilde{f}_n(x) - f(x)\big)^2 dx + \frac{1}{4}\mathrm{Var}\Big(\frac{1}{n}\sum_{j=1}^{n}(X_j^2 - \mathbb{E}(X_1^2))\Big) \int_{\mathbb{R}} f''(x)^2 dx \tag{3.24}$$

where $\mathbb{E}_0$ denotes the expectation with respect to the distribution of n independent random variables distributed from the density f.

Remark 6 The variance $\mathrm{Var}\Big(\frac{1}{n}\sum_{j=1}^{n}(X_j^2 - \mathbb{E}(X_1^2))\Big)$ is also equivalent to $4\,\mathrm{Var}(\frac{1}{n}Y_{n,2})$ (see Ould Haye and Philippe (2003)). Equation (3.24) shows that this term is a ceiling rate of MISE independently of the choice of the kernel and bandwidth.

Proof :

It consists in adapting the proof of Mielniczuk (1997) to the seasonal case. Hereafter, we denote by $\hat{g}$ the Fourier transform of g. Using Hall and Hart (1990) decomposition of the MISE, we have

$$\begin{aligned} \mathrm{MISE}(\tilde{f}_n) = &\int_{\mathbb{R}} \mathbb{E}_0\big(\tilde{f}_n(x) - f(x)\big)^2 \,\mathrm{d}x + \\ &+ \frac{1}{n\pi}\sum_{j=1}^{n-1}(1 - j/n) \int |\hat{K}(bt)|^2 \Big\{\mathrm{Re}\,(\mathbb{E}(e^{it(X_1 - X_{j+1})}) - |\hat{f}(t)|^2\Big\} \,\mathrm{d}t \\ := &\,\mathrm{MISE}_0 + W_n \end{aligned} \tag{3.25}$$

Let f_j be the joint density of (X_1, X_{j+1}). We extend the expansion of f_j obtained by Giraitis et al. (1996) to the order 2 as follows: there exists h_j such that

$$f_j(x, y) = f(x)f(y) + r(j)f'(x)f'(y) + \frac{1}{2}r(h)^2 f''(x)f''(y) + h_j(x, y) \quad \forall (x, y) \in \mathbb{R}^2 \tag{3.26}$$

where r is given in (1.4).

We have

$$\mathbb{E}(e^{it(X_1-X_{j+1})}) = \int e^{it(x-y)} f(x)f(y)\,\mathrm{d}x\,\mathrm{d}y + r(j)\int e^{it(x-y)} f'(x)f'(y)\,\mathrm{d}x\,\mathrm{d}y +$$
$$+\frac{1}{2}r(h)^2\int e^{it(x-y)} f''(x)f''(y)\,\mathrm{d}x\,\mathrm{d}y + \int e^{it(x-y)} h_j(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= |\hat{f}(t)|^2 + r(j)|\widehat{f'}(t)|^2 + \frac{1}{2}r(j)^2|\widehat{f''}(t)|^2 + \widehat{h}_j(t,-t). \tag{3.27}$$

Similarly to Mielniczuk (1997), W_n in (3.25) can be written as

$$W_n = \frac{2}{n}\sum_{j=1}^{n-1}(1-j/n)r(j)\int |K_{m_n}\star f'|^2(t)\,\mathrm{d}t + \frac{1}{n}\sum_{j=1}^{n-1}(1-j/n)r(j)^2\int |K_{m_n}\star f''|^2(t)\,\mathrm{d}t +$$
$$+\frac{1}{n\pi}\sum_{j=1}^{n-1}(1-j/n)\int |\hat{K}(bt)|^2\mathrm{Re}\,\widehat{h}_j(t,-t)\,\mathrm{d}t$$

where $K_{m_n}(x) = m_n^{-1}K(xm_n^{-1})$, and where $f\star g$ is the convolution of f and g. Moreover we have, for $\ell = 1, 2$,

$$\int |K_{m_n}\star f^{(\ell)}|^2(t)\,\mathrm{d}t = \int f^{(\ell)}(t)^2\,\mathrm{d}t + o(1), \quad n\to\infty.$$

We obtain

$$W_n = \frac{2}{n}\sum_{j=1}^{n-1}(1-\frac{j}{n})r(j)\left(\int f'(t)^2\,\mathrm{d}t + o(1)\right) + \frac{1}{n}\sum_{j=1}^{n-1}(1-\frac{j}{n})r(j)^2\left(\int f''(t)^2\,\mathrm{d}t + o(1)\right) +$$
$$+\frac{1}{n\pi}\sum_{j=1}^{n-1}(1-\frac{j}{n})\int |\hat{K}(bt)|^2\mathrm{Re}\,\widehat{h}_j(t,-t)\,\mathrm{d}t. \tag{3.28}$$

According to Giraitis and Surgailis (1990), we have

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{j=1}^{n}(X_j^2 - \mathbb{E}(X_1^2))\Big) = \frac{2}{n^2}\sum_{1\le i,j\le n} r^2(i-j) + O(n^{-1}),$$
$$= \frac{2}{n^2}(nr(0) + 2\sum_{j=1}^{n-1}(n-j)r(j)^2) + O(n^{-1})$$
$$= \frac{4}{n}\sum_{j=1}^{n-1}(1-j/n)r(j)^2 + O(n^{-1}) := \gamma(n) \tag{3.29}$$

Moreover, using the form of r given in (1.4) and the fact that $\alpha < 1/3$, we get

$$\gamma(n) = \frac{4}{n}\sum_{j=1}^{n-1}(1-j/n)j^{-2\alpha}(\sum_{k\in J} a_k\big(\cos j\lambda_k + o(1)\big))^2 + O(n^{-1})$$
$$= \frac{2}{n}\sum_{j=1}^{n-1}(1-j/n)j^{-2\alpha}\sum_{k\in J} a_k^2 + O(n^{-1})$$
$$= \frac{2}{n}n^{1-2\alpha}\left(\frac{1}{1-2\alpha} - \frac{1}{2-2\alpha}\right)\sum_{k\in J} a_k^2 + O(n^{-1})$$
$$= n^{-2\alpha}\frac{1}{(1-2\alpha)(1-\alpha)}\sum_{k\in J} a_k^2 + O(n^{-1}) \sim Cn^{-2\alpha} \tag{3.30}$$

As $\alpha < (1 \wedge \alpha_0)/3$ and using (2.12), we get

$$\frac{2}{n}\sum_{j=1}^{n-1}(1-j/n)r(j) = \frac{1}{n^2}\operatorname{Var}(Y_{n,1}) - r(0)n^{-1} = O(-\alpha_0) + O(n^{-1}) = o(n^{-2\alpha}). \tag{3.31}$$

From (3.28), (3.29), (3.30) and (3.31), we get

$$W_n = \frac{1}{4}\operatorname{Var}\Big(\frac{1}{n}\sum_{j=1}^{n}(X_j^2 - \mathbb{E}(X_1^2))\Big)\int f''(t)^2\,\mathrm{d}t + o(n^{-2\alpha}) + \\ + \frac{1}{n\pi}\sum_{j=1}^{n-1}(1-j/n)\int |\hat{K}(bt)|^2 \operatorname{Re}\widehat{h_j}(t,-t)\,\mathrm{d}t.$$

Since $r(j)^2$ behaves asymptotically as $j^{-2\alpha}$, and

$$\frac{1}{n\pi}\sum_{j=1}^{n-1}(1-j/n)\int |\hat{K}(bt)|^2 \operatorname{Re}\widehat{h_j}(t,-t)\,\mathrm{d}t \le \frac{1}{n\pi}\sum_{j=1}^{n-1}(1-j/n)\int |\widehat{h_j}(t,-t)\,\mathrm{d}t| \tag{3.32}$$

the proof is completed using the following lemma proven below.

Lemma 3.1 *Under the same assumption of Theorem 3.1,*

$$\int |\widehat{h_j}(t,-t)|\,dt = O(j^{-2\alpha-\epsilon}). \tag{3.33}$$

for ϵ an arbitrary positive number smaller than $\dfrac{1-3\alpha}{10}$.

□

Proof of Lemma 3.1 By definition of h_j in (3.26), we have

$$\widehat{h_j}(x,y) = \widehat{f_j}(x,y) - \widehat{f}(x)\widehat{f}(y)(1 - xyr(j) + \frac{1}{2}x^2y^2r(j)^2)$$

We split the integral

$$\int_{\mathbb{R}} |\widehat{h_j}(t,-t)|\,\mathrm{d}t = \int_{|t|>j^{\epsilon}} |\widehat{h_j}(t,-t)|\,\mathrm{d}t + \int_{|t|<j^{\epsilon}} |\widehat{h_j}(t,-t)|\,\mathrm{d}t \tag{3.34}$$

where ϵ is an arbitrary positive number smaller than $\dfrac{1-3\alpha}{10}$.

Under assumption (3.23), Giraitis et al. (1996) proved for the regular long memory that for arbitrary k

$$|\widehat{f_j}(x_1,x_2)| \le c(k)(1+|x|)^{-k}$$

for all $x = (x_1,x_2) \in \mathbb{R}^2$ and

$$|\widehat{f}(x)| \le c(k)(1+|x|)^{-k}$$

for all $x \in \mathbb{R}$.

Their proof can be adapted to the seasonal case i.e. when the coefficients $(b_j)_{j\in\mathbb{N}}$ satisfies (1.3). Using their notation, it suffices to construct a finite set J_1 such that for all $j \in J_1$: $|b_{-j}| > 2|b_{t-j}| + c_1$ where c_1 does not depend on t. Since $(|b_j|)_{j\in\mathbb{Z}}$ is not summable, there exists a subsequence $(j_u)_{u\in\mathbb{Z}}$ such that $b_{-j_u} \neq 0$. We can take J_1 a subset of $\{j_u : u \in \mathbb{Z}\}$ with $[\delta|J_1|] = k+3$. Indeed, for $j \in J_1$, we have $|b_{-j}| > C(J_1)|j^{-(\alpha+1)/2}|$, and for t large enough there exists $\tilde{c}_1$

$$|j|^{-(1+\alpha)/2} > 2/C(J_1)|t-j|^{-(1+\alpha)/2} + \tilde{c}_1.$$

Therefore, there exists c_1 such that for all $j \in J_1$,

$$|b_{-j}| > 2|b_{t-j}| + c_1.$$

For all k', the first integral in (3.34) satisfies

$$\int_{|t|>j^\epsilon} |\widehat{h_j}(t,-t)|\,\mathrm{d}t \le j^{-\epsilon k'} \int_{|t|>j^\epsilon} |t|^{k'} |\widehat{h_j}(t,-t)|\,\mathrm{d}t = O(j^{-\epsilon k'}).$$

Therefore we can take any arbitrary k' such that $k' > (2\alpha+\epsilon)/\epsilon$.

For the second integral in (3.34), it is enough to show that

$$\sup_{|u|<j^\epsilon} |\widehat{h_j}(u)| = O(j^{-2\alpha-2\epsilon}). \tag{3.35}$$

The proof is quite similar to that of equation (2.20) in Giraitis *et al* Giraitis et al. (1996) adding the terms of order two in the expansion.

We write the difference $\widehat{f}_j(x,y) - \widehat{f}(x)\widehat{f}(y)$ from products of the characteristic function ϕ of ξ_1.

$$\begin{aligned}\widehat{f}_j(x,y) - \widehat{f}(x)\widehat{f}(y) &= \prod_{I_1}\prod_{I_1}\prod_{I_1} \phi(xb_{-i} + yb_{t-i}) - \prod_{I_1}\prod_{I_1}\prod_{I_1} \phi(xb_{-i})\phi(yb_{t-i}) := a_1a_2a_3 - a_1'a_2'a_3' \\ &= (a_1' - a_1)a_2a_3 + (a_2' - a_2)a_1'a_3 + (a_3' - a_3)a_1'a_2'\end{aligned}$$

where $I_1 = \{|i| < j^{2\epsilon}\}$, $I_3 = \{|t-i| < j^{2\epsilon}\}$ and $I_3 = \mathbb{Z} - (I_1 \cup I_2)$. We will deduce (3.35) from $|a_i| < 1$, $|a_i'| < 1$ and the following facts, for all $u < t^\epsilon$

$$\begin{aligned} a_i - a_i' &= O(j^{-2\alpha-2\epsilon}, \quad i = 1,2 & (3.36)\\ a_3 - a_3' &= a_3'(-xyr(j) + \frac{1}{2}x^2y^2r(j)^2) + O(j^{-2\alpha-2\epsilon}). & (3.37)\end{aligned}$$

Similarly to Giraitis et al. (1996), we prove (3.36) with $i = 1$ (or similarly for $i = 2$) as follows

$$\begin{aligned}|a_1 - a_1'| &\le \sum_{|i|\le j^{2\epsilon}} |\phi(xb_{-i} + y * b_{j-i}) - \phi(xb_{-i})\phi(y * b_{j-i})| \\ &\le \sum_{|i|\le j^{2\epsilon}} |xb_{-i}|\end{aligned}$$

As $|i| \le j^{2\epsilon}$ and $x \le j^\epsilon$, we have

$$|xb_{-i}| \le Cj^\epsilon j^{-(1+\alpha)/2} = j^{-2\alpha-2\epsilon}O(1)$$

since $\epsilon < \dfrac{1-3\alpha}{10} < \dfrac{1-\alpha/2}{6}$ when $\alpha < 1/3$.

To prove (3.37), we follow the same calculations as Giraitis et al. (1996) page 325. Since $|xb{-}i| + |yb_{j-i}| = o(1)$, we write $a_3 - a_3'$ of the form

$$a_3 - a_3' = a_3'(e^{Q_j(x,y)} - 1) = a_3'(Q_j(x,y) + \frac{1}{2}Q_j(x,y)^2 + o(Q_j(x,y)^2))$$

where

$$\begin{aligned}Q_j(x,y) &= \sum_{i\in I_3} \Psi(xb_{-i}, yb_{j-i}) = -xy\sum_{i\in I_3} b_{-i}b_{j-i} + O(\sum_{i\in I_3}(xb_{-i})^2|yb_{j-i}| + |xb_{-i}||yb_{j-i}|^2) \\ &:= -xy\sum_{i\in I_3} b_{-i}b_{j-i} + R_n\end{aligned}$$

and

$$\Psi(x,y) = \log(\phi(x+y)) - \log(\phi(x)) - \log(\phi(y))$$

and we show that

$$Q_j(x,y) = -xyr(j) + O(\sum_{I_1 \cup I_2} |x||y||b_{-i}||b_{j-i}| + \sum_i x^2|y||b_{-i}|^2|b_{j-i}|) =$$

$$Q_j(x,y)^2 = x^2y^2r(j)^2 + x^2y^2(\sum_{i\in I_1\cup I_2} b_{-i}b_{j-i})^2 - 2x^2y^2 - xy\sum_{i\in\mathbb{Z}} b_{-i}b_{j-i} \sum_{i\in I_1\cup I2} b_{-i}b_{j-i} \\ + R_n^2 - 2R_n xy \sum_{i\in I_3} b_{-i}b_{j-i}$$

For $|x| < j^\epsilon$ et $|y| < j^\epsilon$ we have

$$\sum_{I_1\cup I_2} |x||y||b_{-i}||b_{j-i}| = j^{2\epsilon-(1+\alpha)/2}O(1) = j^{-2\alpha-2\epsilon}O(1)$$

since $\epsilon < (1-3\alpha)/8$ and

$$\sum_i x^2|y||b_{-i}|^2|b_{j-i}| = j^{-\alpha/2-1/2+3\epsilon}O(1) = j^{-2\alpha-2\epsilon}O(1)$$

since $\epsilon < (1-3\alpha)/10$. These asymptotic behaviors ensure that for $|x| < j^\epsilon$ et $|y| < j^\epsilon$ we have

$$a_3 - a_3' = a_3'(xyr(j) + \frac{1}{2}x^2y^2r(j)^2 + O(j^{-2\alpha-2\epsilon})).$$

References

Arteche, J. and Robinson, P. M. (1999). Seasonal and cyclical long memory. In *Asymptotics, nonparametrics, and time series*, volume 158 of *Statist. Textbooks Monogr.*, pages 115–148. Dekker, New York.

Arteche, J. and Robinson, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *J. Time Ser. Anal.*, 21(1):1–25.

Bisognin, C. and Lopes, S. R. C. (2009). Properties of seasonal long memory processes. *Math. Comput. Modelling*, 49(9-10):1837–1851.

Bretagnolle, J. and Huber, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete*, 47(2):119–137.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979)*, volume 757 of *Lecture Notes in Math.*, pages 23–68. Springer, Berlin.

Giraitis, L., Hidalgo, J., and Robinson, P. M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *Annals of Statistics*, 29:987–1023.

Giraitis, L., Koul, H. L., and Surgailis, D. (1996). Asymptotic normality of regression estimators with long memory errors. *Statist. Probab. Lett.*, 29(4):317–335.

Giraitis, L. and Leipus, R. (1995). A generalized fractionally differencing approach in long-memory modeling. *Lietuvos Matematikos Rinkinys*, 35:65–81.

Giraitis, L. and Surgailis, D. (1990). A limit theorem for polynomials of a linear process with long-range dependence. *Lithuanian Mathematical Journal*, 29(2):128–145.

Giraitis, L. and Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *J. Statist. Plann. Inference*, 80(1-2):81–93.

Gray, H. L., Zhang, N.-F., and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, 10:233–257.

Gray, H. L., Zhang, N.-F., and Woodward, W. A. (1994). On generalized fractional processes: a correction. *Journal of Time Series Analysis*, 15:561–562.

Hall, P. and Hart, J. D. (1990). Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Theory Relat. Fields*, 87(2):253–274.

Hassler, U. (1994). (mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis*, 15:19–30.

Hidalgo, J. and Soulier, P. (2004). Estimation of the location and exponent of the spectral singularity of a long memory process. *J. Time Ser. Anal.*, 25(1):55–81.

Ho, H.-C. and Hsing, T. (1996a). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Stat.*, 24(3):992–1024.

Ho, H.-C. and Hsing, T. (1996b). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *J. Appl. Probab.*, 33(1):138–145.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–176.

Hosoya, Y. (1997). A limit theory for long–range dependence and statistical inference on related models. *Annals of Statistics*, 28:105–137.

Hsu, N.-J. and Tsai, H. (2009). Semiparametric estimation for seasonal long-memory time series using generalized exponential models. *J. Statist. Plann. Inference*, 139(6):1992–2009.

Leipus, R. and Viano, M.-C. (2000). Modelling long-memory time series with finite or infinite variance: a general approach. *Journal of Time Series Analysis*, 21(1):61–74.

Major, P. (1981). *Multiple Wiener–Ito Integrals.*, volume 849 of *Lecture Notes in Mathematics*. Springer Verlag, Berlin–Heidelberg–New York.

Mielniczuk, J. (1997). On the asymptotic mean integrated squared error of a kernel density estimator for dependent data. *Statist. Probab. Lett.*, 34(1):53–58.

Ould Haye, M. (2002). Asymptotic behavior of the empirical process for Gaussian data presenting seasonal long-memory. *ESAIM Probab. Statist.*, 6:293–309 (electronic). New directions in time series analysis (Luminy, 2001).

Ould Haye, M. and Philippe, A. (2003). A noncentral limit theorem for the empirical process of linear sequences with seasonal long memory. *Mathematical Methods of Statistics*, 12(3):329–357.

Reisen, V. A., Rodrigues, A. L., and Palma, W. (2006). Estimating seasonal long-memory processes: a Monte Carlo study. *J. Stat. Comput. Simul.*, 76(4):305–316.

Viano, M.-C., Deniau, C., and Oppenheim, G. (1995). Long-range dependence and mixing for discrete time fractional processes. *Journal of Time Series Analysis*, 16(3):323–338.

Whitcher, B. (2004). Wavelet-based estimation for seasonal long-memory processes. *Technometrics*, 46(2):225–238.