Semiparametric Gaussian white noise models in the non-regular case

Ismaël Castillo* and Eric Cator†

December 12, 2007

Abstract

This paper deals with the behavior of some nonlinear semiparametric models in the case where the unknown nuisance function f is not necessarily differentiable. Two models are considered, in the Gaussian white noise framework: estimation of the center of symmetry and estimation of the period of a periodic signal. We obtain the rate of convergence of the sieve maximum likelihood estimators in these models over different functional spaces. In particular, it is shown that if the class controls appropriately the growth to infinity of the Fisher information over the sieve, semiparametric fast rates of convergence are obtained. We also prove a lower bound result which implies that these semiparametric rates are strictly below the parametric ones, meaning there is a significant loss of information, contrary to the regular case.

*Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081HV Amsterdam, Nederland. `i.castillo@few.vu.nl`

†Department of Mathematics, Delft University of Technology, Mekelweg 4, 2628CD Delft, Nederland. `e.a.cator@tudelft.nl`

1 Introduction and main results

The models. Two semiparametric models will be considered in this paper. Let $L^2[0,1]$ be the Hilbert space of all square integrable functions on the interval $[0,1]$. Suppose we observe in continuous time a path X, which is a solution of the diffusion equation

$$dX(t) = f(t/\theta)dt + dW(t) \ , \quad t \in [-T/2, T/2], \quad (T > 0) \tag{1.1}$$

where the unknown function f belongs to $L^2[0,1]$ and is extended 1-periodically, θ is an unknown period parameter which belongs to a subset of $]0, +\infty[$ and W is standard Brownian motion on the interval of observation $[-T/2, T/2]$. We shall work in the asymptotic framework $T \to +\infty$. This model will be referred to as the *period model* throughout the paper.

The second model, which will be referred to as the *translation model* in the sequel, consists of observing a path Y, which is a solution of the diffusion equation

$$dY(t) = f(t-\theta)dt + \varepsilon dW(t) \ , \quad t \in [-1/2, 1/2], \quad (\varepsilon > 0) \tag{1.2}$$

where the unknown function f is *symmetric*, 1-periodic and when restricted to $[0,1]$, belongs to $L^2[0,1]$. The unknown parameter θ, that is the center of symmetry of the *signal* $f(\cdot - \theta)$, is supposed to belong to a compact subset of the real numbers, and W is here standard Brownian motion on $[-1/2, 1/2]$. We shall work in the asymptotic framework $\varepsilon \to 0$.

Context. The aim of the theory of semiparametric models can be described as trying to understand the effect of the unknown nonparametric component of the model in the estimation of a finite-dimensional parameter or functional of interest. The situation is relatively well understood in those models where a certain amount of regularity is present, usually meaning the assumption of local asymptotic normality of the model or, stronger, of differentiability in quadratic mean (see [1] or [11][Chap. 25]), along one-dimensional submodels. Under these conditions, in many cases the effect of the nonparametric nuisance can be understood in terms of scores and tangent sets through the concept of efficient Fisher information, which might be smaller than the Fisher information in the corresponding parametric case. In this case one says that a loss of information occurs. However, even in this case, some gaps remain in the theory, in particular when the efficient Fisher information is zero, where often rates of convergence slower than the parametric ones arise, see for example [9] in a deconvolution framework.

In this article the considered semiparametric models do not necessarily verify the regularity assumptions cited above. In fact, we do not impose a priori any other regularity condition on the function f in model (1.1) than the one necessary for the likelihood function (see (1.3) below) to be defined, that is that f should be square-integrable. To the best of the authors' knowledge, the problem of understanding estimation in non-regular semiparametric contexts has received few attention, and one of the aims of the present article is to shed some light on this issue. In fact, it turns out, as we shall see, that interesting non-parametric effects are observed in the non-regular case which are more or less wiped out by regularity assumptions in the regular case.

Let us now present some already-known results in the Gaussian white noise models (1.1)-(1.2). In the parametric case where f is known, we refer to the book by [7] and the references therein. In the regular parametric case, usual parametric rates are obtained, that is of the order of ε in the translation model (resp. $T^{-3/2}$ in the period model), whereas in the non-regular case, *fast rates* of convergence appear : the more "irregular" the function f, the faster the rate. We shall give more insight in Section 1.2. In the semi-parametric case, in the regular framework, the same rates of convergence (including the constants) as in the parametric case are obtained for models (1.1)-(1.2). We refer to [6] for the period model (1.1) and [4] and references therein for the translation model (1.2). Still in the regular case, non-parametric effects have been shown to occur in the second order terms of the asymptotics, see [4], [5] and [3], where it is shown that semiparametric second order terms are significantly larger than usual parametric ones.

Practical point of view, periodogram and model selection. The problem of estimating the frequency of an unknown periodic signal corrupted by noise is a central issue in signal processing and its applications (telecommunications, laser vibrometry, to name a few), see [8] and the references therein. Models (1.1)-(1.2) can be seen as idealized versions in continuous time of finite sample models. One of the most important methods of estimation involves the maximization of the so-called *cumulated periodogram*. In the context of model (1.1), this is the same as maximizing in τ the quantity

$$\mathrm{L}_K(\tau) \quad = \quad \frac{1}{T}\sum_{|k|\le K}\left|\int_{-T/2}^{T/2}\exp(2ik\pi(t/\tau))dX(t)\right|^2 .$$

One still needs to choose the integer K, which corresponds to the dimension of the finite-dimensional space (or *sieve*) approximating the functional space where f lives. It turns out in practice that the algorithm based on the maximization of $\mathrm{L}_K(\tau)$ is often very sensitive to the choice of K, see [8], where a clever algorithm based on a penalized version of L_K is considered (under, however, regularity assumptions on f). This sensitivity seems not to be foreseen by the theory in the semiparametric *regular* framework, where many choices of K can be shown to produce asymptotically efficient estimators of θ (roughly, if f is regular enough, any K of smaller order than $T^{1/4}$ would do, see [3]). In this article, we shall see that this is no longer true if no strong smoothness assumption is made on the nuisance f: if we allow functions with potentially many significative Fourier coefficients, then we must choose K large enough to recover enough information from the signal. But this also means adding more noise (the contribution of W in the observed X), thus a compromise must be found between these two constraints, which makes the rate of convergence of the estimators based on the cumulated periodogram very sensitive to K, even from a theoretical perspective.

There is another natural reason why considering classes of irregular functions (see below for precise definitions) makes sense in the considered models. In many models, for example partially linear semiparametric models (see for instance [13]), one considers uniform rate results over a convex set of (nuisance) functions (think of Sobolev balls), where in fact the boundary of the set determines the rate, since they are the hardest to estimate. In our case however, it is known already from the parametric case that the more irregular the function, the *faster* the rate. It seems natural, therefore, to try and find uniform results over complements of balls. In that way, it is still the boundary of the set that is hardest to estimate, and therefore determines the rate.

Main results. We address the problem of statistical estimation in models (1.1)-(1.2) in a semiparametric non-regular framework. We establish that uniform fast rates of convergence can be obtained over natural classes of irregular functions by sieve maximum likelihood estimators and determine their rate of convergence. We also establish a lower bound result which implies that the semiparametric rates are strictly below the parametric ones. This lower bound does not quite coincide with the rate of the sieve maximum likelihood estimator and we conjecture that there are semiparametric estimators having a strictly better rate of convergence than sieve maximum likelihood estimators.

1.1 Gaussian white noise models

A special basis of the space $L^2[0,1]$ is the trigonometric basis, denoted $\varepsilon_k(\cdot) = \exp(2i\pi k\cdot)$, for any $k \in \mathbb{Z}$, where $\mathbb{Z}$ denotes the set of all integers. For an element f of $L^2[0,1]$, we denote by $\|\cdot\|$ its L^2-norm and by $\{c_k\}_{k\in\mathbb{Z}}$ the sequence of its Fourier coefficients, so that

$$\|f\|^2 = \int_0^1 f(t)^2 dt, \quad c_k = \int_0^1 f(t)\varepsilon_{-k}(t)dt \quad (k \in \mathbb{Z}).$$

In model (1.2), due to the symmetry of f, we work with

$$f_k = \sqrt{2}\int_0^1 f(t)\cos(2\pi kt)dt \quad (k \geq 0).$$

The only further assumption that we make a priori on the function f is that there exist some real constants ρ, L with $0 < \rho < L$ such that f belongs to the set $\mathcal{F}$ defined as

$$\textbf{(F)} \quad \mathcal{F} \triangleq \mathcal{F}(\rho, L) = \{f \in L^2[0,1], \quad |c_1| \geq \rho, \quad \|f\|^2 \leq L^2\}.$$

We also extend the functions in $\mathcal{F}$ by 1-periodicity so that they can be seen as functions over the real line. Note that we do not impose at this point regularity assumptions on derivatives for f, as in [4], [6], [3] (existence of a bounded second derivative for f in the L^2-sense) or [5] (existence of a bounded $1+\kappa$-derivative, $\kappa > 0$, in the L^2-sense), all the cited work falling in the "regular" framework. The assumption on c_1 is a way to ensure that the energy $\sum_{k\in\mathbb{Z}} k^2|c_k|^2$ of the signal (which might be infinite in our context) is far away from zero. This correspond to the fact that one wants to exclude functions which could be arbitrarily close in L^2 to a constant function, for which estimation of θ in the context of (1.1) or (1.2) is impossible. It also ensures that the function f has 1 as the smallest period, which is important in both models for identifiability reasons.

We further assume that the parameter θ belongs to a set Θ^P (resp. Θ^S) in the period model (resp. translation model) such that

$$\textbf{(T)} \quad \Theta^P = [A, B] \subset]0, +\infty[, \quad \Theta^S = [-\tau_0, \tau_0] \subset]-1/4, 1/4[.$$

In the translation model, one easily checks that to ensure identifiability, the parameter set should have a diameter smaller than 1/2, which explains the choice for Θ^S. As for Θ^P, it would have been possible to choose a varying subset $[A_T, B_T]$ asymptotically covering $]0, +\infty[$, but to avoid technicalities irrelevant to the content of the present paper, we restrict ourselves to the compact case.

The aim of the statistical problem at hand is to understand how and how fast one can estimate the unknown θ, the function f being unknown as well. An estimator will be as usual a measurable function of the observations, which means here a measurable function with respect to the σ-field generated by the process X (respectively Y), the solution of (1.1) (resp. (1.2)), on the Banach space $\mathcal{B}^P$ of continuous functions on $[-T/2, T/2]$ (resp. $\mathcal{B}^S$ on $[-1/2, 1/2]$). We denote by $\mathbf{P}^P_{\theta,f}$ (resp. $\mathbf{P}^S_{\theta,f}$) the probability distribution generated by X (resp. Y) on this space.

The natural statistical object to work with is the *likelihood*, which, in the context of models (1.1) and (1.2), is defined as a likelihood ratio. Let $\mathbf{P}^P_0$ (resp. $\mathbf{P}^S_0$) be the probability distribution generated by Brownian motion on $\mathcal{B}^P$ (resp. by ε times Brownian motion on $\mathcal{B}^S$). Then the likelihoods $L^P(\theta, f)$ and $L^S(\theta, f)$ are defined as the following Radon-Nikodym derivatives, for which explicit expressions are given by Girsanov's formula: for any $\mathbf{X} \in \mathcal{B}^{\mathbf{P}}$ and any $\mathbf{Y} \in \mathcal{B}^{\mathbf{S}}$,

$$\begin{aligned} L^P(\theta, f)(\mathbf{X}) &\triangleq \frac{d\mathbf{P}^P_{\theta,f}}{d\mathbf{P}^P_0}(\mathbf{X}) = \exp\left(\int_{-T/2}^{T/2} f(t/\theta) d\mathbf{X}(t) - \frac{1}{2}\int_{-T/2}^{T/2} f(t/\theta)^2 dt\right), \qquad (1.3) \\ L^S(\theta, f)(\mathbf{Y}) &\triangleq \frac{d\mathbf{P}^S_{\theta,f}}{d\mathbf{P}^S_0}(\mathbf{Y}) = \exp\left(\varepsilon^{-2}\int_{-1/2}^{1/2} f(t-\theta) d\mathbf{Y}(t) - \frac{\varepsilon^{-2}}{2}\int_{-1/2}^{1/2} f(t-\theta)^2 dt\right). \end{aligned}$$

We denote by $\mathbf{E}^P_{\theta,f}$ (resp. $\mathbf{E}^S_{\theta,f}$) the expectation under the probability distribution $\mathbf{P}^P_{\theta,f}$ (resp. $\mathbf{P}^S_{\theta,f}$). In the sequel for simplicity we sometimes drop the index P or S.

1.2 The parametric case

We shall first deal with the parametric non-regular case, where the situation is already well understood (see for instance [7]). In that case, the function f is known and we consider maximum likelihood-type estimators $\overline{\theta}$ of θ. It is intuitively clear that the slower the L^2 difference between the functions with parameters θ and τ goes to zero as $\tau \to \theta$, which corresponds to a more irregular function f, the faster the rate of convergence for $\overline{\theta}$ will be. It turns out that this connection is indeed very direct. We focus here on the translation model, the period case being essentially similar.

The maximum-likelihood estimator is formally defined as

$$\overline{\theta} \triangleq \operatorname*{Argmax}_{\tau \in \Theta^P} \varepsilon^{-2}\int_{-1/2}^{1/2} f(t-\tau) d\mathbf{Y}(t) - \frac{\varepsilon^{-2}}{2}\int_{-1/2}^{1/2} f(t-\tau)^2 dt.$$

Using the fact that the observation process $\mathbf{Y}$ follows (1.2), it is easy to check that

$$\overline{\theta} = \operatorname*{Argmax}_{\tau \in \Theta^P} \int_{-1/2}^{1/2} \{f(t-\tau) - f(t-\theta)\} dW(t) - \frac{1}{2}\int_{-1/2}^{1/2} \{f(t-\tau) - f(t-\theta)\}^2 dt.$$

This leads us to define a class of functions over which the L^2 distance between the signals is controlled from above and below. More precisely, for positives η, $M_1 < M_2$ and $0 < \beta < \alpha < 1$, we define

$$\begin{aligned} \mathcal{F}^S_{\alpha,\beta}(\eta, M_1, M_2) = \{f \in \mathcal{F}: \quad & \forall \tau, \theta \in \Theta^S : |\tau - \theta| < \eta, \\ & \int_{-1/2}^{1/2} \{f(t-\theta) - f(t-\tau)\}^2 dt \geq M_1(\tau-\theta)^{2\alpha} \\ & \int_{-1/2}^{1/2} \{f(t-\theta) - f(t-\tau)\}^2 dt \leq M_2(\tau-\theta)^{2\beta}\}. \end{aligned}$$

The control from below defines the degree of irregularity of the functions in the class, whereas the control from above is to ensure that the supremum of the process $\int_{-1/2}^{1/2}\{f(t-\tau)-f(t-\theta)\}dW(t)$ is essentially controlled by the supremum of the standard deviations. Note that the control from below alone does not guarantee that this last process has continous sample paths.

Note that one can also consider sieve maximum likelihood estimators. Since f is known, so is also $f^{[K]}(\cdot)=\sum_{k=1}^{K} f_k\varepsilon_k(\cdot)$ and we define, for any sequence of integers $K=K(\varepsilon)$ such that $K(\varepsilon)\to+\infty$ as $\varepsilon\to 0$,

$$\overline{\theta}(K) \triangleq \underset{\tau\in\Theta^S}{\text{Argmax}}\ \varepsilon^{-2}\int_{-1/2}^{1/2} f^{[K]}(t-\tau)d\mathbf{Y}(t)-\frac{\varepsilon^{-2}}{2}\int_{-1/2}^{1/2} f^{[K]}(t-\tau)^2dt.$$

As for the sieve maximum likelihood estimator, there is a class of functions particularly adapted to these estimators. For a positive integer K_0, a positive real M and $\alpha\in]0,1[$, we define

$$S_\alpha(K_0,M)=\left\{f\in\mathcal{F}:\quad \forall K\geq K_0:\sum_{k=-K}^{K}k^2|c_k|^2\geq MK^{2-2\alpha}\right\}. \tag{1.4}$$

In the case of model (1.2), the condition is the same as imposing $\sum_{k=1}^{K}k^2f_k^2\geq MK^{2-2\alpha}$. Note that here the process $\int_{-1/2}^{1/2}f^{[K]}(t-\tau)dW(t)$ has sample paths differentiable at any order, and there is no need to control the decay of the tail of Fourier coefficients. Note also that the condition defining the class S_α is stronger than the corresponding lower bound in the definition of the class $\mathcal{F}_{\alpha,\beta}(\eta,M_1,M_2)$, for small enough M_1. This can be seen by expanding the L^2-distance over the Fourier basis and using the bound $|\sin(u)|\geq 2\pi/|u|$ for $0<|u|<\pi$. The parametric rate of convergence is now given by the following theorem.

Theorem 1.1. *Fix positive K_0,M and $\alpha\in]0,1[$. Let us define $l_\varepsilon=\log(\varepsilon^{-1})$ and $\overline{K}_\alpha=\lfloor\{l_\varepsilon\varepsilon\}^{-1/\alpha}\rfloor$. Then for any f in $S_\alpha(K_0,M)$,*

$$\lim_{\varepsilon\to0}\sup_{\theta\in\Theta^S}\mathbf{P}_{\theta,f}\left(|\overline{\theta}(\overline{K}_\alpha)-\theta|\geq l_\varepsilon^{1/\alpha}\varepsilon^{1/\alpha}\right)=0$$

Fix positive η, $M_1<M_2$ and $0<\beta<\alpha<1$. Then for any f in $\mathcal{F}^S_{\alpha,\beta}(\eta,M_1,M_2)$,

$$\lim_{\varepsilon\to0}\sup_{\theta\in\Theta^S}\mathbf{P}_{\theta,f}\left(|\overline{\theta}-\theta|\geq l_\varepsilon\varepsilon^{1/\alpha}\right)=0$$

Notice that arbitrarily high rates can be achieved as the known signal f gets more and more irregular. It will turn out that in the semiparametric setup, this will no longer be possible.

1.3 Semiparametric rates for sieve maximum likelihood estimators

We briefly recall a natural way to obtain estimators in models (1.1)-(1.2). The idea is to use the profile likelihood method (see [11][Chap. 25]) which consists in maximizing the likelihood in two steps: first one maximizes it with respect to the nuisance parameter, obtaining a function -called *profile likelihood*- independent of f, and then one chooses the maximizer of this quantity as estimator. In fact, the first step cannot involve a maximum over the whole space $\mathcal{F}$, which is too large, thus one restricts the maximization to a sieve, which here will be the K-dimensional linear space generated by the first K elements of the trigonometric basis. The final estimator is obtained by choosing K (a choice which turns to be essential) and is called the *sieve maximum likelihood* (or *sieve-MLE*).

This method is particularly useful in the case of models (1.1)-(1.2) since the profile likelihood can be written in a closed form. In the models under consideration, this technique and refinements with weights have been studied in detail in [4] and [3], which is why we will directly give the expression of the obtained profile likelihood in both cases, that is (1.5) for the translation case and (1.11) for the period.

We shall seek for best possible uniform rates of convergence for the sieve-MLE over some class of functions. A first candidate for the class is certainly the whole class $\mathcal{F}$ defined above. However, this class contains (almost) all smooth functions, for which we know that the estimation is in fact more difficult in the considered models and the obtained rate turns out to be ε or $T^{-3/2}$ (the usual parametric $n^{-1/2}$), see Theorems 1.3 and 1.5 and Section 1.4 for a corresponding (and in fact stronger) lower bound result. A smaller class will allow to obtain uniform fast rates for sieve-MLEs : the class $S_\alpha(K_0, M)$ defined by (1.4), which enables to quantify the "irregularity" of a function through a parameter α and which contains all functions at least "α-irregular" in this sense. Note that the definition of $S_\alpha(K_0, M)$ is in fact quite natural in the sense that it gives control over the Fisher information for estimating θ in the sequence (depending on K) of (parametric) models where (θ, f) is of the form $\{(\theta, f_1, \ldots, f_K)\}$. In this sense it suits well the study of a K-dimensional sieve-MLE.

1.3.1 Estimation of the center of symmetry

The profile likelihood method yields the following criterion (see [4]) to maximize

$$l_K(\tau) = \sum_{k=1}^{K} \left| \sqrt{2} \int_{-1/2}^{1/2} \cos(2\pi k(t-\tau)) dY(t) \right|^2, \tag{1.5}$$

where τ belongs to Θ^S, $Y(t)$ is assumed to satisfy (1.2) , and K is an integer to be chosen.

For any $K > 0$, we obtain a sieve-MLE $\hat{\theta}(K)$ by maximizing the profile likelihood

$$\hat{\theta}(K) = \underset{\tau \in \Theta^S}{\text{Argmax}}\ l_K(\tau). \tag{1.6}$$

We are particularly interested in the following choice of K

$$K_S^* = K_{S,\alpha}^* = \begin{cases} \lfloor \varepsilon^{-\frac{2}{1+2\alpha}} \rfloor & \text{if} \quad 1/4 \le \alpha \le 1, \\ \lfloor \varepsilon^{-\frac{4}{1+4\alpha}} \rfloor & \text{if} \quad 0 < \alpha < 1/4 \end{cases}. \tag{1.7}$$

The associated sieve-MLE is defined by

$$\hat{\theta}_S = \underset{\tau \in \Theta^S}{\text{Argmax}}\ l_{K_S^*}(\tau). \tag{1.8}$$

Let us define the rate

$$r_\varepsilon^S = \begin{cases} \varepsilon^{\frac{4}{1+4\alpha}} & \text{if} \quad 0 < \alpha < 1/4, \\ \varepsilon^{\frac{3}{1+2\alpha}} & \text{if} \quad 1/4 \le \alpha \le 1. \end{cases} \tag{1.9}$$

Theorem 1.2. *Let us assume* **(F)** *and* **(T)** *and let* r_ε^S *be defined by* (1.9) *and* $\hat{\theta}_S$ *be defined by* (1.8)*. Then with* $l_\varepsilon = -\log \varepsilon$*, for any* $0 < \alpha < 1$ *and positives* K_0 *and* M*,*

$$\lim_{\varepsilon \to 0} \sup_{\theta \in \Theta^S,\ f \in S_\alpha(K_0, M)} \mathbf{P}_{\theta, f}\left(\left| \hat{\theta}_S - \theta \right| > l_\varepsilon^{1/\alpha} r_\varepsilon^S \right) = 0.$$

The rate of convergence r_ε^S is a *fast rate* in that it is faster than the usual parametric rate $r_{0,\varepsilon}^S = \varepsilon$ (that is $1/\sqrt{n}$ if we take $\varepsilon = 1/\sqrt{n}$). Note that the smaller α (that is the more irregular the function in the sense of the classes $S_\alpha(K_0, M)$), the faster we can estimate θ, as we can see immediately from the expression (1.9). A close examination of the proof of Theorem 1.2 reveals that the rate of $\hat{\theta}(K)$ is quite sensitive to the choice of the cut-off K. Note also that there is a change of slope in the power of the rate (1.9) at $\alpha = 1/4$, which is reminiscent of the nonparametric effect observed in the problem of estimating the L^2-norm, where $\alpha = 1/4$ is also a transition point.

We also consider the sieve-MLE $\tilde{\theta}_S$ for $K = \tilde{K} \triangleq \lfloor \varepsilon^{-1/2} \rfloor$ that is

$$\tilde{\theta}_S = \operatorname*{Argmax}_{\tau \in \Theta^S} l_{\tilde{K}}(\tau), \tag{1.10}$$

and we denote by $r_{0,\varepsilon}^S = \varepsilon$ the rate in the smooth case.

Theorem 1.3. *Let us assume* **(F)** *and* **(T)** *and let* $\tilde{\theta}_S$ *be defined by* (1.10)*. Then with* $l_\varepsilon = -\log \varepsilon$*,*

$$\lim_{\varepsilon \to 0} \sup_{\theta \in \Theta^S,\ f \in \mathcal{F}} \mathbf{P}_{\theta, f}\left(\left| \tilde{\theta}_S - \theta \right| > l_\varepsilon r_{0,\varepsilon}^S \right) = 0.$$

Theorem 1.3 extends the semiparametric results known in the regular case, where f is differentiable and the Fisher information is finite, saying that semiparametric estimation is possible, whatever the regularity of f in $\mathcal{F}$, at the rate ε (up to a log factor).

1.3.2 Estimation of the period

The profile likelihood method leads to the criterion, for any positive integer K,

$$\mathrm{L}_K(\tau) = \frac{1}{T} \sum_{|k| \le K} \left| \int_{-T/2}^{T/2} \varepsilon_k(t/\tau) dX(t) \right|^2, \tag{1.11}$$

where the process X satisfies (1.1) and $\varepsilon_k(\cdot) = \exp(2ik\pi \cdot)$. As already noted in [6], there is an additional difficulty in using (1.11) compared to (1.5), in that periodicity of f implies that $\mathrm{L}_K(2\theta)$ is roughly of the same order as $\mathrm{L}_K(\theta)$ and thus one needs an extra localization argument. We overcome this difficulty by considering the smallest approximate minimizer of (1.11) by defining

$$\mathcal{E}_K = \left\{ \tau \in \Theta^P,\ \ \mathrm{L}_K(\tau) \ge (1 - \delta) \sup_{\tau \in \Theta^P} \mathrm{L}_K(\tau) \right\} \tag{1.12}$$

$$e_K = \inf \mathcal{E}_K, \tag{1.13}$$

where δ should satisfy

$$0 < \delta < \frac{\rho^2}{L^2}.$$

The inequality ensures that the $\tau's$ equal to fractions of θ (see Lemma 2.4) will not be in $\mathcal{E}_K$ with high probability (see the proof of Lemma 2.6).

For any $K > 0$, we obtain an estimator of the period (despite the localization argument involved we still refer to it as sieve-MLE), which, adopting as in [6] the following definition for a ball of center x of radius r

$$B(x,r) = \left\{\tau \in \Theta^P,\ \left|\frac{x}{\tau} - 1\right| < r\right\}, \tag{1.14}$$

is defined as

$$\theta^*(K) = \operatorname*{Argmax}_{\tau \in B(e_K, 1/4)} \mathrm{L}_K(\tau). \tag{1.15}$$

Let us define

$$K_P^* = K_{P,\alpha}^* = \begin{cases} \lfloor T^{\frac{1}{1+2\alpha}} \rfloor & \text{if} \quad 1/4 \leq \alpha \leq 1, \\ \lfloor T^{\frac{2}{1+4\alpha}} \rfloor & \text{if} \quad 0 < \alpha < 1/4 \end{cases}. \tag{1.16}$$

The associated sieve-MLE is defined by

$$\hat{\theta}_P = \theta^*(K_{P,\alpha}^*) = \operatorname*{Argmax}_{\tau \in B(e_{K_{P,\alpha}^*}, 1/4)} \mathrm{L}_{K_{P,\alpha}^*}(\tau) \quad \text{if } 1/4 \leq \alpha \leq 1. \tag{1.17}$$

In fact, similar to the translation model, we have a transition when $\alpha = 1/4$. In the period case, when $0 < \alpha < 1/4$, we define the estimator $\hat{\theta}_P$ as a two-step estimator. First, the estimator $\theta^*(K_{P,1/4}^*)$ yields a first intermediate rate T^{-2} (see Subsection 2.2.4) which is strengthened by considering the final estimator

$$\hat{\theta}_P = \operatorname*{Argmax}_{\tau \in \Theta^P \cap B(\theta^*(K_{P,1/4}^*), T^{-2})} \mathrm{L}_{K_{P,\alpha}^*}(\tau) \quad \text{if } 0 < \alpha < 1/4. \tag{1.18}$$

We also need to restrict slightly the classes $\mathcal{F}$ and $S_\alpha(K_0, M)$. We assume that there exist $\beta > 0$ and $L_1 > L$ such that f belongs to the following class

$$\mathcal{F}^+ = \mathcal{F}(\rho, L) \cap \left\{f = \{c_k\},\quad \sum_{k\in\mathbb{Z}} k^{2\beta}|c_k|^2 \leq L_1^2\right\}.$$

Similarly for any $\alpha > \beta$ we define $S_{\alpha,\beta}(K_0, M) = S_\alpha(K_0, M) \cap \mathcal{F}^+$.

Let us define

$$r_T^P = \begin{cases} T^{-\frac{3+4\alpha}{1+4\alpha}} & \text{if} \quad 0 < \alpha < 1/4 \\ T^{-\frac{5+4\alpha}{2+4\alpha}} & \text{if} \quad 1/4 \leq \alpha \leq 1. \end{cases} \tag{1.19}$$

Theorem 1.4. *Let r_T^P be defined by (1.19) and let $\hat{\theta}_P$ be defined by (1.17), (1.18). Then, for any $0 < \beta < \alpha < 1$ and any positive K_0, M, with $l_T = -\log T$,*

$$\lim_{T\to+\infty} \sup_{\theta\in\Theta^P,\ f\in S_{\alpha,\beta}(K_0,M)} \mathbf{P}_{\theta,f}\left(\left|\hat{\theta}_P - \theta\right| > l_T^{1/\alpha} r_T^P\right) = 0.$$

This result is slightly more difficult and technical than the corresponding result in the translation case, but the rate is essentially the same, once one replaces ε in (1.9) by T^{-1} and multiplies the result by T^{-1}. Note that we require slightly more here than the function being in L^2 by considering the class $\mathcal{F}^+$. This condition appears when one wants to establish uniform convergence to zero of some remainder terms in the profile likelihood criterion. These terms are in fact zero in the translation case due to orthogonality arguments (see also the proof of Theorem 1.4). However the condition is not very strong since it suffices that a very small β exists for the results to hold.

As with the translation model, we can complement the previous result by considering the sieve-MLE $\tilde{\theta}_P$ for $K = \tilde{K}_P \triangleq \lfloor T^{1/3} \rfloor$, that is

$$\tilde{\theta}_P = \underset{\tau \in \Theta^P}{\text{Argmax}} \; \mathrm{L}_{\tilde{K}_P}(\tau). \tag{1.20}$$

and we denote by $r_{0,T}^P = T^{-3/2}$ the rate in the smooth case.

Theorem 1.5. *Let $\tilde{\theta}_P$ be defined by (1.20). Then with $l_T = -\log T$,*

$$\lim_{T \to +\infty} \sup_{\theta \in \Theta^P, \; f \in \mathcal{F}^+} \mathbf{P}_{\theta,f}\left(\left| \tilde{\theta}_P - \theta \right| > l_T r_{0,T}^P \right) = 0.$$

Theorem 1.5 extends the results of [6] to any f in $\mathcal{F}^+$, whereas Theorem 1.4 asserts than the smooth rate $r_{0,T}^P = T^{-3/2}$ can be notably improved when working uniformly over the slightly smaller class $S_{\alpha,\beta}(K_0, M)$.

1.4 Lower bounds

As we have seen in Section 1.2, the sieve-MLE attains faster uniform rates over the class $S_\alpha(K_0, M)$ than the smooth rates. It is then natural to investigate the best possible rate of convergence of semiparametric estimators in the minimax sense over such classes. In particular, we would like to know whether it is possible or not to do as well as in the parametric case, which is the case in the regular framework, at least asymptotically.
Let us define

$$\underline{r_T} = T^{-\frac{7+8\alpha}{2+8\alpha}} \qquad \text{and} \qquad \underline{r_\varepsilon} = \varepsilon^{\frac{5}{1+4\alpha}}. \tag{1.21}$$

Theorem 1.6. *For any $0 < \beta < \alpha < 1$, positive K_0, M, there exist C, C' such that, using definition (1.21),*

$$\liminf_{T \to +\infty} \inf_{\hat{\theta}} \sup_{\theta \in \Theta^P, \; f \in S_{\alpha,\beta}(K_0,M)} \mathbf{E}_{\theta,f}^P((\hat{\theta} - \theta)^2 \underline{r_T}^{-2}) \geq C > 0$$

$$\liminf_{\varepsilon \to 0} \inf_{\hat{\theta}} \sup_{\theta \in \Theta^S, \; f \in S_\alpha(K_0,M)} \mathbf{E}_{\theta,f}^S((\hat{\theta} - \theta)^2 \underline{r_\varepsilon}^{-2}) \geq C' > 0,$$

where the infima are taken over all estimators of θ in the period (resp. translation) model.

Theorem 1.6 states that the semiparametric rate is slower than the rate obtained in Section 1.2 in the parametric case over $S_{\alpha,\beta}(K_0, M)$. This means that a significant loss of information occurs. Note also that this lower bound is actually strictly faster than the rates we found for the sieve-MLE.

Another interesting question is the following: what happens if we seek for uniform results over classes larger than S_α? Generalisations of these classes are the following. Let us define, for positives η, M_1 and $0 < \alpha < 1$,

$$\begin{aligned}
\mathcal{F}_\alpha^P(\eta, M_1) &= \{f \in \mathcal{F}, \quad \forall\, \tau, \theta \in \Theta^P : |\tau - \theta| < \eta, \\
&\qquad \int_{-T/2}^{T/2} \{f(t/\theta) - f(t/\tau)\}^2\, dt \geq M_1(\tau - \theta)^{2\alpha}\}, \\
\mathcal{F}_\alpha^S(\eta, M_1) &= \{f \in \mathcal{F}, \quad \forall\, \tau, \theta \in \Theta^S : |\tau - \theta| < \eta, \\
&\qquad \int_{-T/2}^{T/2} \{f(t - \theta) - f(t - \tau)\}^2\, dt \geq M_1(\tau - \theta)^{2\alpha}\}.
\end{aligned}$$

Classes of this type are quite natural in that they define the irregularity using the intrinsic distance on the statistical problem, that is here the L^2-distance between the signals. However, if one seeks for fast uniform rates, they are, somehow, too large.

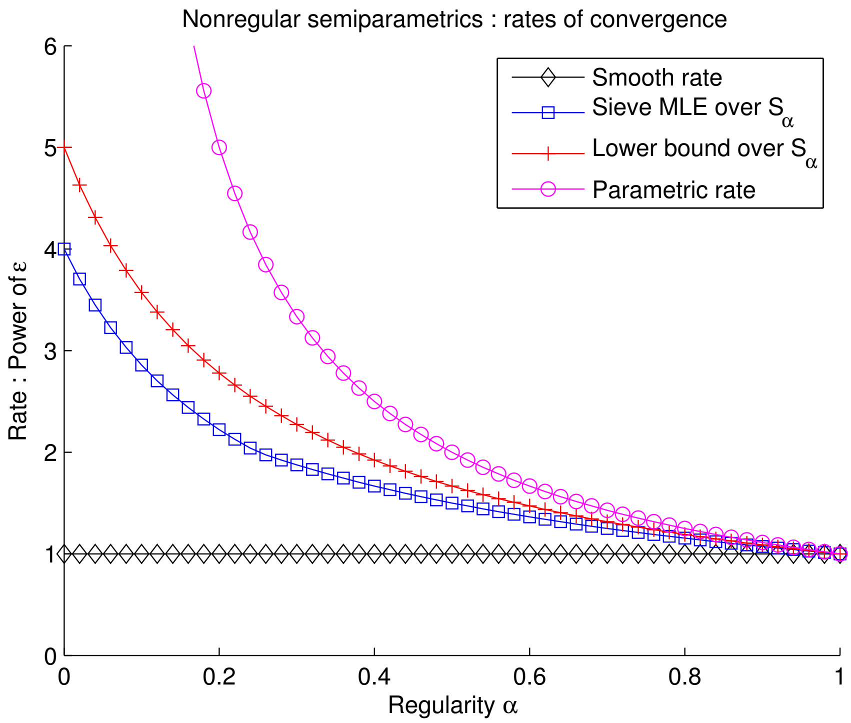
Theorem 1.7. *For any $0 < \alpha < 1$ and positive η, M_1, we can choose L, defining the class $\mathcal{F}(\rho, L)$, big enough such that there exist C, C' with*

$$\begin{aligned}
&\liminf_{T \to +\infty} \inf_{\hat{\theta}} \sup_{\theta \in \Theta^P,\ f \in \mathcal{F}_\alpha^P(\eta, M_1)} \mathbf{E}_{\theta,f}^P((\hat{\theta} - \theta)^2 T^3) \geq C > 0 \\
&\liminf_{\varepsilon \to 0} \inf_{\hat{\theta}} \sup_{\theta \in \Theta^S,\ f \in \mathcal{F}_\alpha^S(\eta, M_1)} \mathbf{E}_{\theta,f}^S((\hat{\theta} - \theta)^2 \varepsilon^{-2}) \geq C' > 0,
\end{aligned}$$

where the infima are taken over all estimators of θ in the period (resp. translation) model.

In the semiparametric context, it is thus not possible to obtain a better minimax rate of convergence over the classes $\mathcal{F}_\alpha$ than in the case where f is smooth. The fact that we might need to choose L very big, is related to the fact that the upper bound on the L^2-norm in $\mathcal{F}$ might conflict with the lower bound on the L^2 distance of the signals.

1.5 Discussion and perspectives



The obtained rates of convergence in the translation case are summarized in the figure above. The main message is that uniform fast rates of convergence are obtained for sieve-MLE estimators (the 'square'-curve) over appropriate spaces and that the best possible rates in the minimax sense over these spaces are strictly below the parametric ones ('plus' and 'circle'-curves). This loss of speed is even worse if one considers larger spaces ('diamond'-curve).

Now there are two questions, both linked to the "gap" between the 'square' and 'plus'-curves. The first is : can we do better with sieve-MLEs (1.6) than the rate we obtain for the choice (1.7) of K^*? In other words, is the choice $K = K^*$ optimal? Although we shall not prove it here, we think it is, see the discussion at the end of the proof of Theorem 1.2.

The second is, if the sieve-MLEs cannot do better than the 'square curve', is the 'plus' curve not optimal or can we improve estimation? We conjecture that there are estimators outperforming the sieve-MLEs.

2 Proofs : upper bounds

Along the proofs, we use a number of self-contained Lemmas whose proofs are given in Section 4. We abbreviate the distributions $\mathbf{P}^P_{\theta,f}$ and $\mathbf{P}^S_{\theta,f}$ by $\mathbf{P}_{\theta,f}$ or even $\mathbf{P}$, since it will always be clear from the context to which model we refer to. Let us denote, for a process Y indexed by a subset Θ of the real numbers,

$$\| Y \| \triangleq \sup_{t\in\Theta} Y(t).$$

The notation $\lesssim$ (resp. $\gtrsim$) is used for "smaller (resp. larger) than or equal to a universal constant times". Sometimes we make universal constants explicit and denote them by C.

Furthermore, $\mathcal{O}$ and o are the usual Landau symbols. For any integers m, n, we denote by $m \wedge n$ their minimum.

2.1 Semi-parametric case: center of symmetry

The criterion l_K defined in (1.5) can be written

$$l_K(\tau) = \sum_{k=1}^{K} f_k^2 \cos^2(2\pi k(\tau-\theta)) + \varepsilon\eta_1(\tau) + \varepsilon^2\eta_2(\tau),$$

where

$$\eta_1(\tau) = 2\sum_{k=1}^{K} f_k \left\{\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*\right\} \cos(2\pi k(\tau-\theta)) \tag{2.1}$$

$$\eta_2(\tau) = \sum_{k=1}^{K} \left\{\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*\right\}^2, \tag{2.2}$$

and

$$\xi_k = \sqrt{2}\int_{-1/2}^{1/2} \cos(2\pi kt)dW(t)$$

$$\xi_k^* = \sqrt{2}\int_{-1/2}^{1/2} \sin(2\pi kt)dW(t).$$

Note that $\{\xi_k\}_{k\geq 1}$ and $\{\xi_k^*\}_{k\geq 1}$ form two independent sequences of independent $\mathcal{N}(0,1)$ variables. Note also that η_2 is a pure noise term (that is, it does not depend on f). Let us write

$$l_K(\tau) - l_K(\theta) = -\sum_{k=1}^{K} f_k^2 \sin^2(2\pi k(\tau-\theta)) + \varepsilon(\eta_1(\tau)-\eta_1(\theta)) + \varepsilon^2(\eta_2(\tau)-\eta_2(\theta)).$$

Using the properties of the Fourier basis and setting $\delta = \tau - \theta$ one checks that

$$\mathbf{E}(\{\eta_1(\tau)-\eta_1(\theta)\}^2) = 4\sum_{k=1}^{K} f_k^2 \sin^2(2\pi k\delta) \triangleq 4v(\delta, K), \tag{2.3}$$

$$\mathbf{E}(\{\eta_2(\tau)-\eta_2(\theta)\}^2) = 4\sum_{k=1}^{K} \sin^2(2\pi k\delta). \tag{2.4}$$

2.1.1 Control of processes

To prove Theorem 1.2, it is necessary to be able to control the processes $\eta_1(\tau) - \eta_1(\theta)$ and $\eta_2(\tau) - \eta_2(\theta)$. For both, we establish that the supremum of the process over the sets of interest is controlled up to a logarithmic factor by the supremum of the standard deviations. For Gaussian processes (thus a priori only for $\eta_1(\tau) - \eta_1(\theta)$), this property is known to be true if one is able to control the entropy of the indexing set of the process with respect to the natural semimetric given by its covariance structure (see for example [10] Theorem 2.4 or [12] Appendix A.2.2). Here we shall instead use the fact that the two processes at stake are differentiable a.s. (which comes from their explicit dependence in τ, see Equations (2.1) and (2.2)) and we will apply Lemma 4.1. The apparent drawback of this rougher method is that a logarithmic factor appears. However, we were not able to get rid of this log-factor when using the more refined techniques mentioned above.

Lemma 2.1. *There exist positive constants C_1, C_2, C_3, C_4 such that for K large enough, for any positive real y,*

$$\begin{aligned} \mathbf{P}\left(\|\eta_2 - K\| > y\sqrt{K}\right) &\leq C_1 K \exp(-C_2 y\{y \wedge K^{1/2}\}) \\ \mathbf{P}\left(\|\eta_2'\| > yK^{3/2}\right) &\leq C_3 K^2 \exp(-C_4 y\{y \wedge K^{3/2}\}). \end{aligned}$$

In particular, note that Lemma 2.1 implies that if K is polynomial in ε^{-1}, then $\mathbf{P}(\|\eta_2 - K\| > l_\varepsilon\sqrt{K})$ and $\mathbf{P}(\|\eta_2'\| > l_\varepsilon K^{3/2})$ decrease to zero faster than any polynomial in ε. Notice that the same holds for $\mathbf{P}(\|\eta_2(\cdot) - \eta_2(\theta)\| > l_\varepsilon\sqrt{K})$.

Lemma 2.2. *Let γ_ε be a sequence tending to zero as $\varepsilon \to 0$ and $v(\cdot, K)$ be defined by (2.3). There exist positive constants C_5, C_6 such that for K large enough, for any positive real y,*

$$\sup_{\theta\in\Theta^S, f\in\mathcal{F}} \mathbf{P}_{\theta,f}\left(\sup_{\gamma_\varepsilon\leq|\tau-\theta|<2\tau_0} \frac{|\eta_1(\tau) - \eta_1(\theta)|}{v(\tau-\theta, K)^{1/2}} \geq y\right) \leq C_5\gamma_\varepsilon^{-4}K^4\exp(-C_6 y^2).$$

The proof of these lemmas can be found in Section 4.

2.1.2 Proof of Theorem 1.2

The starting point of the proof is quite standard. One has to study the criterion $l_K(\tau)$ and proving that it takes its global maximum close to $\tau = \theta$ gives us already a first "rough" rate of convergence. Then, studying its local behavior around $\tau = \theta$ by analyzing the dependence of $l_K(\tau) - l_K(\theta)$ in $\tau - \theta$ hopefully enables us to sharpen the rate.

The "standard" approach in the smooth case where f has some derivatives is to expand the criterion around $\tau = \theta$ using Taylor's formula and to control the rest terms by bounds on the derivatives of l_K, see for instance [4], [5] (and [6], [3] in the case of the period model). This approach cannot be used here, at least not until a sufficiently fast rate has been reached. The reason is that as long as $\delta > K^{-1}$, the functional $v(\delta, \varepsilon)$ does not behave like δ^2 but like $\delta^{2\alpha}$. A similar phenomenon (but this time independently of f) occurs with the pure noise part $\eta_2(\tau) - \eta_2(\theta)$: we shall use the bound $|\sin(2\pi k\delta)| \leq 1 \wedge 2\pi k|\delta|$, which yields different estimates depending on how close τ is to θ. In the "smooth" case, this discussion was unnecessary since K could be chosen small enough, corresponding to the fact that the smoother the function, the smaller the number of significative Fourier coefficients.

We establish the rate of convergence in several steps. In the proof, to simplify the notation, the cut-off K_S^* defined by (1.7) will simply be denoted by K. Let $\gamma_\varepsilon \geq \gamma_\varepsilon'$ be two sequences of positive real numbers and $v(\delta, K)$ defined by (2.3), then

$$\begin{aligned}
\mathbf{P}(\gamma'_\varepsilon < |\hat{\theta}_S - \theta| \leq \gamma_\varepsilon) &\leq \mathbf{P}\left(\sup_{\tau:\ \gamma'_\varepsilon < |\tau - \theta| \leq \gamma_\varepsilon} l(\tau) \geq l(\theta) \right) \\
&\leq \mathbf{P}\left(\sup_{\delta:\ \gamma'_\varepsilon < |\delta| \leq \gamma_\varepsilon} -v(\delta, K) + 2\varepsilon(\eta_1(\tau) - \eta_1(\theta)) + \varepsilon^2(\eta_2(\tau) - \eta_2(\theta)) \geq 0 \right) \\
&\leq \mathbf{P}\left(4\varepsilon \sup_{\gamma'_\varepsilon < |\delta| \leq \gamma_\varepsilon} \frac{\eta_1(\tau) - \eta_1(\theta)}{v(\delta, K)^{1/2}} \geq \inf_{\gamma'_\varepsilon < |\delta| \leq \gamma_\varepsilon} v(\delta, K)^{1/2} \right) \qquad (2.5) \\
&\quad + \mathbf{P}\left(2\varepsilon^2 \sup_{\gamma'_\varepsilon < |\delta| \leq \gamma_\varepsilon} \frac{\eta_2(\tau) - \eta_2(\theta)}{v(\delta, K)^{1/2}} \geq \inf_{\gamma'_\varepsilon < |\delta| \leq \gamma_\varepsilon} v(\delta, K)^{1/2} \right) \qquad (2.6)
\end{aligned}$$

We first establish that the estimator $\hat{\theta}_S$ is at least convergent at the rate $\gamma_{1,\varepsilon} = l_\varepsilon \varepsilon K^{1/4}$. For this we prove that the right-hand side of the last display tends to zero if $\gamma_\varepsilon = 2\tau_0$ and $\gamma'_\varepsilon = \gamma_{1,\varepsilon}$. Note that this last quantity tends to zero due to the choice $K = K^*$.

Note that $v(\delta, K) \geq f_1^2 \sin^2(2\pi\delta)$. Since $|\delta| < 2\tau_0$ and $2\tau_0$ is strictly less that $1/2$, we have that $\inf_{\delta:\ |\delta| > \gamma_{1,\varepsilon}} v(\delta, K)^{1/2} \gtrsim |f_1|\gamma_{1,\varepsilon} \gtrsim \gamma_{1,\varepsilon}$ using **(F)**. Thus using Lemma 2.2, we obtain that (2.5) tends to zero as $\varepsilon \to 0$. To see that the same holds for (2.6), note that for some small $C > 0$,

$$\begin{aligned}
(2.6) \quad &\leq \mathbf{P}\left(\varepsilon^2 \sup_{\delta:\ \gamma'_\varepsilon < |\delta| \leq \gamma_\varepsilon} \eta_2(\tau) - \eta_2(\theta) \geq \inf_{\delta:\ |\delta| > \gamma_\varepsilon} v(\delta, K) \right) \qquad (2.7) \\
&\leq \mathbf{P}\left(\sup_{\delta:\ \gamma'_\varepsilon < |\delta| \leq \gamma_\varepsilon} \eta_2(\tau) - \eta_2(\theta) \geq C l_\varepsilon^2 K^{1/2} \right).
\end{aligned}$$

Finally we use Lemma 2.1.

Now we shall use repeatedly (2.5)-(2.6), improving each time on the rate of convergence γ'_ε assuming that f belongs to the class $S_\alpha(K_0, M)$.

Let us check that the rate of convergence over $S_\alpha(K_0, M)$ is at least $\gamma_{2,\varepsilon} = D_2 x_\varepsilon K^{-1}$, where $x_\varepsilon = l_\varepsilon^{1/4\alpha}$ and D_2 is a large enough constant (we do this only in the case $\alpha \leq 3/4$, since for $3/4 < \alpha \leq 1$, $\gamma_{1,\varepsilon} \leq \gamma_{2,\varepsilon}$). Thus in this paragraph we work on the set $\{\delta : \gamma_{2,\varepsilon} < |\delta| \leq \gamma_{1,\varepsilon}\}$. Since δ now tends to zero, $\lfloor 1/4|\delta| \rfloor \geq 1/8|\delta|$. Using the fact that for any $x \in [-\pi/2, \pi/2]$, $|\sin(x)| \geq 2|x|/\pi$, we have that for δ positive,

$$v(\delta, K) \geq \sum_{k=1}^{K \wedge \lfloor 1/4\delta \rfloor} f_k^2 \sin^2(2\pi k\delta) \gtrsim \delta^2 \sum_{k=1}^{K \wedge 1/8\delta} k^2 f_k^2.$$

The definition of the class S_α now implies that $v(\delta, K) \gtrsim \delta^2\{K \wedge 1/8\delta\}^{2-2\alpha} \gtrsim \delta^{2\alpha}$ and thus $\inf_{\delta:\ \gamma_{2,\varepsilon} < |\delta| \leq \gamma_{1,\varepsilon}} v(\delta, K) \gtrsim x_\varepsilon^{2\alpha} K^{-2\alpha}$. To show that (2.5) and (2.6) (via (2.7)) tend to zero, we make use of Lemmas 2.2, 2.1. This can be done if the two following conditions are satisfied:

$$\varepsilon^{-1} x_\varepsilon^\alpha K^{-\alpha} \gtrsim \sqrt{l_\varepsilon} \text{ and } \varepsilon^{-2} x_\varepsilon^{2\alpha} K^{-2\alpha} \gtrsim \sqrt{l_\varepsilon K}.$$

This happens if K is less than $K_{max} = \varepsilon^{\frac{-4\alpha}{1+4\alpha}}$. Since this is the case for the choice $K = K^*$, we have proved that $\hat{\theta}_S$ achieves the rate $\gamma_{2,\varepsilon}$.

Finally we check that $\hat{\theta}_S$ achieves the rate $\gamma_{3,\varepsilon} = x_\varepsilon^4 r_\varepsilon^S$. Note that on $\{\delta :\ \gamma_{3,\varepsilon} < |\delta| \leq \gamma_{2,\varepsilon}\}$ we have that

$$v(\delta, K) \geq \sum_{k=1}^{K/x_\varepsilon} f_k^2 \sin^2(2\pi k\delta) \gtrsim \delta^2 (K/x_\varepsilon)^{2-2\alpha},$$

and thus

$$\inf_{\delta:\ \gamma_{3,\varepsilon}<|\delta|\leq\gamma_{2,\varepsilon}} v(\delta, K)^{1/2} \quad \gtrsim \quad \gamma_{3,\varepsilon}(K/x_\varepsilon)^{1-\alpha},$$

which in view of the definitions of K^* and r_ε^S together with Lemma 2.2 implies that (2.5) tends to zero. Now note that since η_2 is differentiable, there exists a (random) c_τ such that

$$\eta_2(\tau) - \eta_2(\theta) = (\tau - \theta)\eta_2'(c_\tau) = \delta\eta_2'(c_\tau).$$

Thus

$$\sup_{\delta:\ \gamma_{3,\varepsilon}<|\delta|\leq\gamma_{2,\varepsilon}} \frac{|\eta_2(\tau) - \eta_2(\theta)|}{v(\delta, K)^{1/2}} \lesssim \sup_{\delta:\ \gamma_{3,\varepsilon}<|\delta|\leq\gamma_{2,\varepsilon}} \frac{|\delta \sup_{\tau\in\Theta} \eta_2'(\tau)|}{|\delta|(K/x_\varepsilon)^{1-\alpha}} \lesssim (K/x_\varepsilon)^{\alpha-1} \sup_{\tau\in\Theta} |\eta_2'(\tau)|.$$

We finally obtain that

$$\begin{aligned}(2.6) \quad &\leq \quad \mathbf{P}\left(\sup_{\tau\in\Theta} |\eta_2'(\tau)| \gtrsim \varepsilon^{-2}(K/x_\varepsilon)^{1-\alpha} \inf_{\delta:\ \gamma_{3,\varepsilon}<|\delta|\leq\gamma_{2,\varepsilon}} v(\delta, K)^{1/2}\right) \\ &\leq \quad \mathbf{P}\left(\sup_{\tau\in\Theta} |\eta_2'(\tau)| \gtrsim \varepsilon^{-2}(K/x_\varepsilon)^{2-2\alpha}\gamma_{3,\varepsilon}\right).\end{aligned}$$

Using the definitions of K^* and r_ε^S we see that $\varepsilon^{-2}(K/x_\varepsilon)^{2-2\alpha}\gamma_{3,\varepsilon} \gtrsim K^{3/2}\sqrt{l_\varepsilon}$. Now Lemma 2.1 implies that (2.6) tends to zero, which concludes the proof. □

Note that the rate of convergence is determined by the value of $\tau-\theta$ such that the pure-noise process $\varepsilon^2(\eta_2(\tau)-\eta_2(\theta))$ becomes dominant in the balancing of $-\sum_{k=1}^K f_k^2 \sin^2(2\pi k(\tau-\theta)) + \varepsilon(\eta_1(\tau) - \eta_1(\theta)) + \varepsilon^2(\eta_2(\tau) - \eta_2(\theta))$. Values of K essentially faster or slower that K^* eventually lead to rates for $\tau - \theta$ slower than r_ε^S when doing the preceding balancing.

2.1.3 Proof of Theorem 1.3

We reproduce the beginning of the proof of Theorem 1.2 and obtain a rate $\gamma_\varepsilon = l_\varepsilon \varepsilon \tilde{K}^{1/4}$ (note that this intermediate rate is faster than $\tilde{K}^{-1}$). Now we proceed as in the last step of the proof for $\hat{\theta}_S$. Using the estimates $v(\delta, \tilde{K}) \gtrsim \delta^2$ together with Lemmas 2.2 and 2.1, we obtain the rate of convergence $l_\varepsilon \varepsilon$, which concludes the proof of the Theorem.

Remark. Note that the process which limits the rate of convergence in Theorem 1.2 is always η_2 whereas η_1 determines the rate in Theorem 1.3. In other words, considering small values of K makes η_2 negligible.

2.1.4 Parametric case : Proof of Theorem 1.1

The proof for the sieve estimator $\overline{\theta}(\overline{K})$ is very much in the spirit of the proof of Theorem 1.2 but much easier since there are just two terms in the criterion and no pure noise term η_2. Indeed, starting again from the definition of $\overline{\theta}(\overline{K}_\alpha)$ and using the noteworthy fact that the functions $f^{[K]}(\cdot - \tau)$ and $(f - f^{[K]})(\cdot - \theta)$ are orthogonal in $L^2[-1/2, 1/2]$, one obtains

$$\begin{aligned}\overline{\theta}(K) &= \operatorname*{Argmax}_{\tau \in \Theta^S} \int_{-1/2}^{1/2} \{f^{[K]}(t-\tau) - f^{[K]}(t-\theta)\} dW(t) \\ &\quad -\frac{1}{2}\int_{-1/2}^{1/2} \{f^{[K]}(t-\tau) - f^{[K]}(t-\theta)\}^2 dt.\end{aligned}$$

The remaider of the proof now closely follows the proof of Theorem 1.2. The deterministic part in the preceding display equals $-2\sum_{k=1}^K f_k^2 \sin^2(\pi k \delta)$ whereas the process part plays the role of $\eta_1(\tau) - \eta_1(\theta)$.

The proof for the maximum likelihood estimator $\overline{\theta}$ is slightly different, since one has to deal with the function f itself and not $f^{[K]}$. But the assumption that f belongs to the class $\mathcal{F}^S_{\alpha,\beta}(\eta, M_1, M_2)$ enables to control the entropy of the indexing set of the process $\tau \to \int_{-1/2}^{1/2} \{f(t-\tau) - f(t-\theta)\} dW(t)$ and eventually to control the supremum of the process with the supremum of the standard deviations. The remainder of the prooof is as before and is omitted. □

2.2 Semi-parametric case: estimation of the period

The profile likelihood method leads to the criterion $\mathrm{L}_K(\tau)$ which can be written as

$$\begin{aligned}\mathrm{L}_K(\tau) &= \frac{1}{T}\sum_{|k|\le K}\left|\int_{-T/2}^{T/2} e^{2ik\pi t/\tau} dX(t)\right|^2, \\ &= \Gamma(\tau) + \zeta(\tau) + \Psi(\tau),\end{aligned}$$

where, with the notation $\varepsilon_k(\cdot) = \exp(2ik\pi\cdot)$, and omitting the dependence in K,

$$\Gamma(\tau) = \frac{1}{T}\sum_{|k|\le K}\left|\int_{-T/2}^{T/2} \varepsilon_k(t/\tau) f(t/\theta) dt\right|^2 \tag{2.8}$$

$$\zeta(\tau) = \frac{2}{T}\sum_{|k|\le K}\int_{-T/2}^{T/2} \varepsilon_k(t/\tau) f(t/\theta) dt \int_{-T/2}^{T/2} \varepsilon_k(t/\tau) dW(t) \tag{2.9}$$

$$\Psi(\tau) = \frac{1}{T}\sum_{|k|\le K}\left|\int_{-T/2}^{T/2} \varepsilon_k(t/\tau) dW(t)\right|^2. \tag{2.10}$$

Let us define a local parameter λ by $\lambda \triangleq \tau^{-1} - \theta^{-1}$ and let

$$\mathcal{V}(\lambda, K) \triangleq T\sum_{|k|\le K} |c_k|^2 \left(k^2\lambda^2 T^2 \wedge 1\right). \tag{2.11}$$

This last quantity is the analog in the period model of $v(\delta, K)$ defined by (2.3) in the translation case.

Let us highlight the main differences with Theorem 1.2. The key ingredients to control the process parts ζ, Ψ are Lemma 4.1 and an eigenvalues-type decomposition for Ψ (see Lemma 4.4). But the major difficulty with respect to the translation case lies in fact in the study and behavior of the deterministic part Γ of the criterion L. For any f in $\mathcal{F}^+$, we first show (Lemmas 2.3, 2.4) that Γ takes its maximum close to θ or one of its multiples. This enables us to get a preliminary rate of roughly $1/T$ for the sieve-MLE (Lemma 2.6). Local expansions around $\tau = \theta$ uncover the functional $\mathcal{V}(\lambda, K)$ in Lemma 2.7, at least for K of smaller order than T. For larger values of K, an extra localization step is necessary. Once the appropriate local expansions are obtained, the proof follows quite closely the one of Theorem 1.2.

In Subsections 2.2.1,2.2.2 and 2.2.3, K (unless otherwise stated) is any sequence of positive integers such that, as T tends to $+\infty$, $K \to +\infty$ and $K = o(T^2)$. Regularity-dependent results and precise choices of K are given in Subsection 2.2.4. In the sequel, the Landau notation $o, \mathcal{O}$, are meant in a *uniform sense* with respect to τ and f. More precisely, when we write that '$G(\tau, f) = H(\tau, f)(1 + o(1))$ holds for $\tau \in \mathcal{T}$ and $f \in \mathcal{G}$', we mean $\sup_{\tau \in \mathcal{T},\ f \in \mathcal{G}} |G(\tau, f)/H(\tau, f) - 1| = o(1)$.

2.2.1 Profile likelihood, global behavior

For any real number x, let us denote by $\Delta(x)$ its distance to $\mathbb{Z}$ and by $]x[$ the (smallest) integer realizing this distance.

Lemma 2.3. *For any $f \in L^2[0, 1]$ and any $\tau \in \Theta^P$, as T tends to $+\infty$,*

$$\Gamma(\tau) = G(\tau) + \|f\|^2 \mathcal{O}\left(\log K + \frac{\log^2 K}{T}\right) \tag{2.12}$$

where

$$G(\tau) = T \sum_{|k| \le K} |c_{]k\theta/\tau[}|^2 \left|\hat{\phi}\left(\frac{T}{\theta}\Delta(k\theta/\tau)\right)\right|^2 \tag{2.13}$$

For any $p, j \in \mathbb{Z}$, we denote by $gcd(p, j)$ the greatest common divisor of p and j.

Lemma 2.4. *Let ε_T and γ_T be positive real-valued sequences such that $\gamma_T \to +\infty$ and*

$$T\varepsilon_T \to +\infty, \quad \gamma_T T^{-1} \to 0, \quad \gamma_T \varepsilon_T \to 0.$$

Then for any $f \in \mathcal{F}^+$ and any $\tau \in \Theta^P$, as T tends to $+\infty$,

$$\begin{aligned}
G(\tau) &= \{|c_0|^2 + o(1)\}T \quad && \text{if } \tau \notin \cup_{j \ge 1,\ 0 < p \le \gamma_T} B(j\theta/p, \varepsilon_T) && (2.14)\\
G(\tau) &\le \{\textstyle\sum_{|q| \le \lfloor \frac{K}{j} \rfloor} |c_{qp}|^2 + o(1)\}T \quad && \text{if } \tau \in \cup_{0 < p \le \gamma_T,\ gcd(p,j)=1} B(j\theta/p, \varepsilon_T) && (2.15)\\
G(\theta) &= \{\textstyle\sum_{|k| \le K} |c_k|^2\}T. && && (2.16)
\end{aligned}$$

The supremum of the process parts ζ and Ψ is controlled with the following lemma.

Lemma 2.5. *For any integer p, as $T \to +\infty$,*

$$\sup_{\theta \in \Theta^P,\, f \in \mathcal{F}^+} \mathbf{P}_{\theta,f} \left(\| \Psi - K \| > l_T^2 \sqrt{K} \right) = o(T^{-p}) \tag{2.17}$$

$$\sup_{\theta \in \Theta^P,\, f \in \mathcal{F}^+} \mathbf{P}_{\theta,f} \left(\| \zeta \| > l_T \sqrt{T} \right) = o(T^{-p}). \tag{2.18}$$

The proof of these lemmas can be found in Section 4.

2.2.2 Sieve-MLE : consistency when $K = o(T)$

The results of the preceding subsection allow us to derive a first rough rate of convergence for $\theta^*(K)$ if $K = o(T)$ as $T \to +\infty$.

Lemma 2.6. *For any integer p and any K such that $K = o(T)$, as $T \to +\infty$,*

$$\sup_{\theta \in \Theta^P, f \in \mathcal{F}^+} \mathbf{P}_{\theta,f} \left(\theta^*(K) \notin B\left(\theta, \frac{l_T}{T}\right) \right) = o(T^{-p}).$$

Proof. Let us define the event

$$\mathcal{A}_0 = \left\{ \sup_{\tau \in \Theta^P} |\zeta(\tau) + \Psi(\tau) - K| \leq l_T^2 T^{1/2} \right\}. \tag{2.19}$$

The complement of $\mathcal{A}_0$ has probability $o(T^{-p})$ for any integer p, thanks to Lemma 2.5.

It suffices to show that for T large enough, for any $\omega \in \mathcal{A}_0$, we have $\theta^*(\omega) \in B(\theta, \varepsilon_T)$.

Let us use Lemmas 2.3 and 2.4 with $\varepsilon_T = l_T/T$ and $\gamma_T = T^{1/2}$. On the event $\mathcal{A}_0$,

$$\sup_{\tau \in \Theta^P} \mathrm{L}(\tau) \leq T \left\{ \sum_{|k| \leq K} |c_k|^2 + o(1) + C l_T^2 T^{-1/2} \right\} \leq T \left\{ \sum_{|k| \leq K} |c_k|^2 + o(1) \right\}.$$

Similarly, on $\mathcal{A}_0$, due to Lemma 2.3,

$$\mathrm{L}(\theta) \geq T \left\{ \sum_{|k| \leq K} |c_k|^2 + o(1) \right\}.$$

Thus on $\mathcal{A}_0$,

$$\frac{\mathrm{L}(\theta)}{\sup_{\tau \in \Theta^P} \mathrm{L}(\tau)} \geq 1 + o(1),$$

which means that for T large enough, $\theta \in \mathcal{E}_T$ on $\mathcal{A}_0$.

On the other hand, by a similar argument using Lemma 2.4, one checks that on $\mathcal{A}_0$,

$$\mathcal{E}_T \subset \cup_{j \geq 1,\ 0 < p \leq \gamma_T} B(j\theta/p, \varepsilon_T).$$

If p is not equal to 1 and τ belongs to $B(j\theta/p, \varepsilon_T)$, we first note that the convergence of $\sum_{k=1}^{K} |c_k|^2$ to $\|f\|^2$ and of $\sum_{|q| \leq \lfloor K/j \rfloor} |c_{qp}|^2$ to $\sum_{q \in \mathbb{Z}} |c_{qp}|^2$ are uniform over functions f in

$\mathcal{F}^+$. Indeed, for any f in $\mathcal{F}^+$, $\sum_{|k|>K}|c_k|^2 \leq K^{-2\beta}L_1^2$. Now use that $|c_1| \geq \rho$ and $\|f\| \leq L$ to see that

$$\sum_{q\in\mathbb{Z}}|c_{qp}|^2 \leq \left(1-\frac{\rho^2}{L^2}\right)\sum_{k\in\mathbb{Z}}|c_k|^2.$$

This easily implies that, on $\mathcal{A}_0$, uniformly over $\mathcal{F}^+$, for T large enough, the values of $\mathrm{L}(\tau)$ for τ in $B(j\theta/p,\varepsilon_T)$ and $p>1$ remain far away from the supremum of L, that is, on $\mathcal{A}_0$,

$$\mathcal{E}_T \subset \cup_{j\geq 1}B(j\theta,\varepsilon_T).$$

Thus using the above intermediate result stating that $\theta \in \mathcal{E}_T$, we obtain that on $\mathcal{A}_0$,

$$e_T \in B(\theta,\varepsilon_T).$$

Noting that $B(e_T,1/4)\cap B(2\theta,\varepsilon_T)=\emptyset$, the definition of θ^* implies that on $\mathcal{A}_0$,

$$\theta^* \in B(\theta,\varepsilon_T).$$

□

2.2.3 Profile likelihood, local behavior

This subsection investigates the local behavior of the different parts of the criterion L. Once one is sufficiently close to θ, which means here if $|\tau^{-1}-\theta^{-1}|=|\lambda|<K^{-1}$, the situation very much resembles the translation case. The differences $\Gamma(\tau)-\Gamma(\theta)$ and $\zeta(\tau)-\zeta(\theta)$ are very much related to the functional $\mathcal{V}(\lambda,K)$ (see (2.11)), while $\Psi(\tau)-\Psi(\theta)$ is controlled using a "derivatives" method through the process $\mathcal{R}$ defined by

$$\mathcal{R}(\tau) \triangleq \frac{\Psi(\tau)-\Psi(\theta)}{(\tau-\theta)} \text{ if } \tau\neq\theta, \quad \mathcal{R}(\theta)=\Psi'(\theta). \tag{2.20}$$

Lemma 2.7. *There exist positive c, C such that, uniformly over $\tau\in\Theta^P$ such that $|\tau^{-1}-\theta^{-1}|=|\lambda|<K^{-1}$,*

$$-c\mathcal{V}(\lambda,K)+\mathcal{O}(l_T) \leq \Gamma(\tau)-\Gamma(\theta) \leq -C\mathcal{V}(\lambda,K)+\mathcal{O}(l_T).$$

Lemma 2.8. *There exists a constant $C>0$ such that, for any $\tau\in\Theta^P$ with $|\tau^{-1}-\theta^{-1}|=|\lambda|<K^{-1}$, the variance of the centered Gaussian process $\zeta(\tau)-\zeta(\theta)$ satisfies*

$$\mathbf{E}_{\theta,f}\left(\{\zeta(\tau)-\zeta(\theta)\}^2\right) \leq C\mathcal{V}(\lambda,K). \tag{2.21}$$

Moreover, if $\mathcal{R}$ is defined by (2.20), then for any integer p, as T tends to $+\infty$,

$$\mathbf{P}\left(\sup_{\tau\in\Theta^P}|\mathcal{R}(\tau)|>TK^{3/2}l_T\right)=o(T^{-p}). \tag{2.22}$$

2.2.4 Proof of Theorem 1.4

By Lemma 2.6, we already have a rate of convergence of $l_T T^{-1}$. As for Theorem 1.2, we shall improve step by step on this rate.

The case $1/4\leq\alpha\leq 1$. Let us choose $K=K^*_{P,\alpha}$, see Equation 1.16. Note that since for this choice of K we have $K = o(l_T^{-1}T)$, the rate for λ we obtain is such that $\lambda = o(K^{-1})$, thus the local expansions provided by Lemmas 2.7 and 2.8 can be used. This enables us to get analogue versions of Lemmas 2.1 and 2.2 in the case of the period model. The precise statement are omitted. Then, we reproduce, using Lemmas 2.7 and 2.5, the proof of the corresponding part of Theorem 1.2 (that is the one to obtain the rate $\gamma_{2,\varepsilon}$) and we obtain that for any K_0, M, as T tends to $+\infty$,

$$\sup_{\theta\in\Theta^P, f\in S_\alpha(K_0,M)} \mathbf{P}_{\theta,f}\left(\theta_P^* \notin B\left(\theta, \frac{1}{KT}\right)\right) \to 0.$$

Once this rate is obtained, the last step of the proof is to show that one obtains the rate $l_T r_T^P$. Using Lemmas 2.7 and 2.8, this follows closely the corresponding part of the proof of Theorem 1.2, and therefore we omit the rest of the proof and leave details to the interested reader.

The case $0 < \alpha < 1/4$. One cannot directly make use of Lemma 2.7 for the estimator $\theta^*_{P,\alpha}$ since $K = K^*_{P,\alpha}$ is larger than T. But we can make use of the estimator $\theta^*(K^*_{P,1/4})$ to get a rate T^{-2} (use Theorem 1.4 in the case $\alpha = 1/4$, for which the theorem has been established above). Note that for any $0 < \alpha < 1/4$, this rate is faster than $1/K^*_{P,\alpha}$. Thus defining the estimator $\hat{\theta}_P$ as in (1.18), we obtain that its rate is at least $T^{-2} = o(1/K^*_{P,\alpha})$ as $T \to +\infty$ and Lemmas 2.7 and 2.8 can now be used. The remainder of the proof is as in the case $1/4 \le \alpha \le 1$ and is omitted.

The proof of Theorem 1.5 is similar to the translation case and is left to the reader.

3 Proofs: Lower bound results

To prove a lower bound on the rate of convergence when estimating θ (either the location of symmetry or the period, in our case), we will follow a well-known approach, which is outlined beautifully in Pollard's so far unpublished book Asymptopia: choose $\mathbf{P}_{\theta,g_1}, \ldots, \mathbf{P}_{\theta,g_m}$ and $\mathbf{P}_{\tau,g_1}, \ldots, \mathbf{P}_{\tau,g_m}$ with τ far enough away from θ, such that

$$\mathbf{Q}_\theta \triangleq \frac{1}{m}\sum_{k=1}^m \mathbf{P}_{\theta,g_k} \quad \text{and} \quad \mathbf{Q}_\tau \triangleq \frac{1}{m}\sum_{k=1}^m \tilde{\mathbf{P}}_{\tau,g_k}$$

remain close in L^1. Then we can use the following inequality:

$$\inf_{\hat{\theta}} \sup_{(\theta,f)} \mathbf{E}_{\theta,f}\left(\hat{\theta} - \theta\right)^2 \ge \frac{1}{4}(\theta-\tau)^2(1 - \frac{1}{2}\|\mathbf{Q}_\theta - \mathbf{Q}_\tau\|_1).$$

So our goal is to maximize $(\theta-\tau)^2$, keeping $\|\mathbf{Q}_\theta - \mathbf{Q}_\tau\|_1$ away from 2. In both the proofs of Theorem 1.7 and of Theorem 1.6, we do this by choosing a smooth function f_m (which may in fact not depend on m), and bounding respectively $\chi^2(\mathbf{P}_{\theta,f_m}, \mathbf{Q}_\theta)$, $\chi^2(\mathbf{P}_{\tau,f_m}, \mathbf{Q}_\tau)$ (these can be bounded using the same technique) and $\chi^2(\mathbf{P}_{\theta,f_m}, \mathbf{P}_{\tau,f_m})$, which is relatively easy since f_m is smooth. Since for any two probability measures $\mathbf{P}$ and $\mathbf{Q}$ we have

$$\|\mathbf{P} - \mathbf{Q}\|_1 \le \sqrt{\chi^2(\mathbf{P}, \mathbf{Q})},$$

we can then use the triangle inequality for the L^1-norm to bound $\|\mathbf{Q}_\theta - \mathbf{Q}_\tau\|_1$. This scheme is depicted in Figure 1.

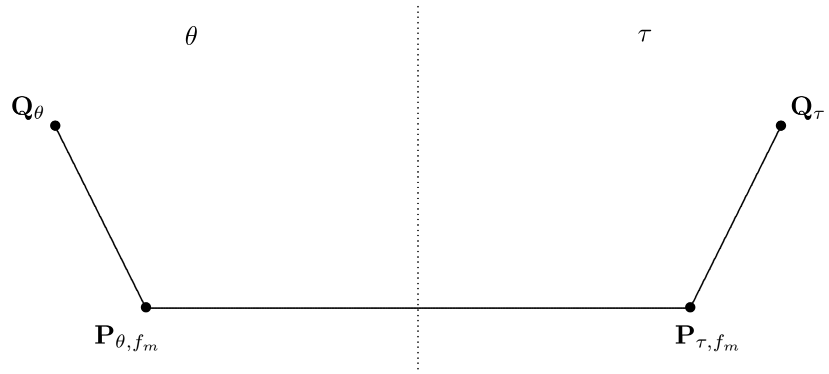


Figure 1: Idea behind lower bound proofs.

3.1 Proof of Theorem 1.7

To prove Theorem 1.7 in the translation case, we need the following lemma.

Lemma 3.1. *Suppose $f \in L^2[0,1]$ and the pertubations $\eta_1, \dots, \eta_m \in L^2[0,1]$. Define*

$$\mathbf{P}_1 = \mathbf{P}_{\theta,f} \quad \textit{and} \quad \mathbf{P}_2 = \frac{1}{m}\sum_{k=1}^{m} \mathbf{P}_{\theta,f+\eta_k}.$$

Then in the translation model, we have

$$\chi^2(\mathbf{P}_1, \mathbf{P}_2) + 1 = \mathbf{E}_{\mathbf{P}_1}\left(\frac{d\mathbf{P}_2}{d\mathbf{P}_1}\right)^2 = \frac{1}{m^2}\sum_{1\le i,j\le m} \exp\left(\frac{1}{\varepsilon^2}\int_{-1/2}^{1/2}\{\eta_i\eta_j\}(t-\theta)dt\right),$$

and in the period model we have

$$\chi^2(\mathbf{P}_1, \mathbf{P}_2) + 1 = \mathbf{E}_{\mathbf{P}_1}\left(\frac{d\mathbf{P}_2}{d\mathbf{P}_1}\right)^2 = \frac{1}{m^2}\sum_{1\le i,j\le m} \exp\left(\int_{-T/2}^{T/2}\{\eta_i\eta_j\}(t/\theta)dt\right).$$

The proof of this Lemma follows from the Girsanov formula and some elementary calculations. Note that the result does not depend on the function f.

Now choose the smooth function

$$f(x) = \cos(2\pi x).$$

We will choose pertubations $\eta_1, \dots, \eta_m$ such that $f + \eta_k \in \mathcal{F}_\alpha^S(\eta, M_1)$, and such that $\mathbf{P}_{\theta,f}$ and $\frac{1}{m}\sum_{k=1}^{m} \mathbf{P}_{\theta,f+\eta_k}$ are arbitrarily close together as $m \to \infty$. This will show that you cannot estimate θ in the class $\mathcal{F}_\alpha^S$ faster than in the smooth case.

We start with an element $\phi \in L^2[-1/2,1/2]$, with support in $(-1/4,1/4)$, and such that $\phi \in \mathcal{F}_\alpha^S(1/2, R)$, for some constant R we will specify later. Recall that elements of $\mathcal{F}$ are extended by 1-periodicity. However, to define our perturbations, we shall just need the values of ϕ within $(-1/2,1/2]$, so we set $\phi_0(x) = \phi(x)\mathbf{1}_{(-1/2,1/2]}(x)$ for all x, where $\mathbf{1}_B$ is the indicator of B. For fixed m and any x in $(-1/2,1/2]$, we define for $k = 1, \dots, m$

$$\eta_k(x) = \sqrt{m}\,\phi_0\left(m\left(x + \frac{1}{4m} - \frac{k}{2m}\right)\right) + \sqrt{m}\,\phi_0\left(m\left(-x + \frac{1}{4m} - \frac{k}{2m}\right)\right).$$

Then we define $\eta_k(x)$ for any x in $\mathbb{R}$ by 1-periodicity. These rescaled, translated and symmetrized versions of ϕ are all orthogonal, since their supports are disjoint. We wish to check that $f + \eta_k \in \mathcal{F}_\alpha^S(\eta, M_1)$. First we remark that bounding the first Fourier coefficient of $f + \eta_k$ from below is not a problem, since

$$\left| \int_{-1/2}^{1/2} \eta_k(x) e^{-2\pi i x} \, dx \right| \leq \int_{-1/2}^{1/2} |\eta_k(x)| \, dx \leq \frac{2}{\sqrt{m}} \int_{-1/2}^{1/2} |\phi(x)| \, dx,$$

so this can be done by choosing m large enough. Also note that for every $m \geq 2$ and every $k = 1, \ldots, m$, we have

$$\|\eta_k\|^2 = 2\|\phi\|^2,$$

so the L^2-norm of the η_k's remains bounded. Finally, using the inequality $(a+b)^2 \geq \frac{1}{2}a^2 - b^2$, we see that

$$\begin{aligned}
&\int_{-1/2}^{1/2} [f(x-\theta) + \eta_k(x-\theta) - f(x-\tau) - \eta_k(x-\tau)]^2 \, dx \\
&\qquad \geq \frac{1}{2} \int_{-1/2}^{1/2} [\eta_k(x-\theta) - \eta_k(x-\tau)]^2 \, dx - \int_{-1/2}^{1/2} [f(x-\theta) - f(x-\tau)]^2 \, dx \\
&\qquad \geq \frac{1}{2} \int_{-1/2}^{1/2} [\eta_k(x-\theta) - \eta_k(x-\tau)]^2 \, dx - 2\pi^2(\theta-\tau)^2.
\end{aligned}$$

In order to control the first part of the right-hand side, we assume that m is even, since otherwise $\eta_{(m+1)/2}$ will have period $1/2$, which might cause problems. If m is even and $|\theta - \tau| > 1/2m$, but $|\theta - \tau| < \eta \leq 1/2$, then for every k, $\eta_k(x - \theta)$ will have at least one "bump" (either the left scaled copy of ϕ or the right one) which is disjoint from the support of $\eta_k(x - \tau)$, and vice versa. This means that

$$\int_{-1/2}^{1/2} [\eta_k(x-\theta) - \eta_k(x-\tau)]^2 \, dx \geq 2\|\phi\|^2 \qquad (\text{for } |\theta - \tau| > \frac{1}{2m}).$$

Furthermore, if $|\theta - \tau| \leq 1/2m$, then it is not hard to see that

$$\begin{aligned}
\int_{-1/2}^{1/2} [\eta_k(x-\theta) - \eta_k(x-\tau)]^2 \, dx &\geq \int_{-1/2}^{1/2} [\phi(y - m\theta) - \phi(y - m\tau)]^2 \, dy \\
&\geq Rm^{2\alpha}(\theta - \tau)^{2\alpha}.
\end{aligned}$$

To guarantee that $f + \eta_k \in \mathcal{F}_\alpha^S(\eta, M_1)$, it is therefore enough to make sure that $\|\phi\|$ is larger than some lower bound depending on η and M_1 (but not on m!), in which case the relevant inequalities are satisfied for all m big enough. Increasing $\|\phi\| = \|\eta_k\|$ might cause the function $f + \eta_k$ to leave the class $\mathcal{F}(\rho, L)$, unless we choose L big enough. Now define

$$\mathbf{Q}_\theta = \frac{1}{m} \sum_{k=1}^{m} \mathbf{P}_{\theta, f+\eta_k}.$$

Then Lemma 3.1 tells us that, using the fact that the η_k's are orthogonal,

$$\|\mathbf{P}_{\theta,f} - \mathbf{Q}_\theta\|_1 \leq \sqrt{\chi^2(\mathbf{P}_{\theta,f}, \mathbf{Q}_\theta)} \leq \frac{1}{\sqrt{m}} \sqrt{e^{2\|\phi\|_2^2/\varepsilon^2} - 1}.$$

This means that by choosing m large enough, we can put $\mathbf{P}_{0,\theta}$ arbitrarily close to $\mathbf{P}_{\theta,f}$. Furthermore, it is clear that

$$\|\mathbf{P}_{\theta,f} - \mathbf{Q}_\theta\|_1 = \|\mathbf{P}_{\tau,f} - \mathbf{Q}_\tau\|_1.$$

Since f is smooth, we can keep $\|\mathbf{P}_{\theta,f} - \mathbf{P}_{\tau,f}\| \leq 1$ when $|\theta - \tau|$ is of the order ε:

$$\begin{aligned}
\|\mathbf{P}_{\theta,f} - \mathbf{P}_{\tau,f}\| &\leq \sqrt{\chi^2(\mathbf{P}_{\theta,f}, \mathbf{P}_{\tau,f})} \\
&= \sqrt{\exp\left(\varepsilon^{-2}\int_0^1 (f(t-\tau) - f(t-\theta))^2\,dt\right) - 1} \\
&\leq 1,
\end{aligned}$$

where the last inequality holds if

$$\int_0^1 (f(t-\tau) - f(t-\theta))^2\,dt = 2\sin^2(2\pi(\theta-\tau)) \leq \frac{1}{2}\varepsilon^2.$$

This can be done by choosing $|\theta - \tau| = \varepsilon/4\pi$. By the arguments belonging to Figure 1 the minimax result of Theorem 1.7 then follows. The proof for the period case is slightly more technical, but it uses the same pertubations and ideas (of course the smooth rate in this case is $T^{-3/2}$).

3.2 Proof of Theorem 1.6

Here we will need slightly more complicated perturbations than in the previous section. We therefore also need a new lemma:

Lemma 3.2. *Suppose $\phi_1, \ldots, \phi_m \in L^2[0,1]$ are orthogonal and $f \in L^2[0,1]$ and $\|\phi_i\|_2^2 \leq M$. Extend these functions periodically. Suppose the interval I has integer length. Define $\mathbf{P}_f$ as the measure corresponding to the model*

$$dX(t) = f(t)dt + dW(t) \quad (t \in I).$$

Define for $m \geq 1$ $\mathcal{W} = \{-1,1\}^m$ and for $w \in \mathcal{W}$, define

$$\phi_w = \sum_{i=1}^m w_i \phi_i.$$

Finally, define

$$\mathbf{Q}^{\mathcal{W}} = 2^{-m} \sum_{w \in \mathcal{W}} \mathbf{P}_{f+\phi_w}.$$

Then there exists a constant $C > 0$ depending only on M such that

$$\mathbf{E}_{\mathbf{P}_f}\left(\frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f}\right)^2 \leq \exp\left(C\sum_{i=1}^m \left(\int_I \phi_i^2(t)\,dt\right)^2\right).$$

Proof. We start by noting that if F is a primitive of f, we have

$$\begin{aligned} \frac{d\mathbf{P}_{f+\phi_w}}{d\mathbf{P}_f}(X) &= \exp\left(\int_I \phi_w(t)\,dX(t) - \frac{1}{2}\int_I \phi_w^2(t)\,dt - \int_I f(t)\phi_w(t)\,dt\right) \\ &= \exp\left(\int_I \phi_w(t)\,d(X-F)(t) - \frac{1}{2}\int_I \phi_w^2(t)\,dt\right). \end{aligned}$$

This means that, using the fact that the ϕ_i's are orthogonal,

$$\begin{aligned} \frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f}(X) &= 2^{-m}\sum_{w\in\mathcal{W}} \exp\left(\int_I \phi_w(t)\,d(X-F)(t) - \frac{1}{2}\int_I \phi_w^2(t)\,dt\right) \\ &= \exp\left(-\frac{1}{2}\sum_{i=1}^n \int_I \phi_i^2(t)\,dt\right) 2^{-m}\sum_{w\in\mathcal{W}} \exp\left(\int_I \phi_w(t)\,d(X-F)(t)\right) \\ &= \exp\left(-\frac{1}{2}\sum_{i=1}^n \int_I \phi_i^2(t)\,dt\right) \times \\ &\quad \prod_{i=1}^m \frac{1}{2}\left[\exp\left(\int_I \phi_i(t)\,d(X-F)(t)\right) + \exp\left(-\int_I \phi_i(t)\,d(X-F)(t)\right)\right]. \end{aligned}$$

So we get, using the fact that the random variables $\{\int_I \phi_i(t)\,dW(t)\}$ are independent,

$$\begin{aligned} \mathbf{E}_{\mathbf{P}_f}\left(\frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f}\right)^2 &= \exp\left(-\sum_{i=1}^n \int_I \phi_i^2(t)\,dt\right) \times \\ &\quad \mathbf{E}\left(\prod_{i=1}^m \frac{1}{4}\left[2 + \exp\left(2\int_I \phi_i(t)\,dW(t)\right) + \exp\left(-2\int_I \phi_i(t)\,dW(t)\right)\right]\right) \\ &= \prod_{i=1}^m \frac{1}{2}\left[\exp\left(-\int_I \phi_i^2(t)\,dt\right) + \exp\left(\int_I \phi_i^2(t)\,dt\right)\right] \end{aligned}$$

Now choose $C > 0$ such that for all $|u| \leq M$, $\cosh(u) \leq 1 + Cu^2 \leq e^{Cu^2}$. Then we get

$$\mathbf{E}_{\mathbf{P}_f}\left(\frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f}\right)^2 \leq \exp\left(C\sum_{i=1}^m \left(\int_I \phi_i^2(t)\,dt\right)^2\right).$$

□

The above lemma can be used for both the translation and the period case. The proof for both cases is again very similar, and we will concentrate on the translation case. Remember the definition of our class:

$$S_\alpha(K_0, M) = \{f \in \mathcal{F} : \quad \forall K \geq K_0 : \sum_{k=-K}^{K} k^2|c_k|^2 \geq MK^{2-2\alpha}\}.$$

Choose $C > 1$ big enough such that $g \in S_\alpha(K_0, M)$, where g has Fourier coefficients

$$g_k = Ck^{-\frac{1}{2}-\alpha}\ (k \geq 1) \text{ and } g_{-k} = g_k.$$

Now fix $m \geq 2$ and denote $\hat{\psi}(k) = \int_0^1 \psi(x) e^{-2ik\pi x} dx$ the Fourier coefficients of a square-integrable function ψ. Define $\phi_1, \ldots, \phi_m \in L^2[0,1]$ by their Fourier coefficients.

$$\begin{aligned} \hat{\phi}_1(m) &= C \cdot m^{-\frac{1}{2}-\alpha} \\ \hat{\phi}_i(k) &= C \cdot k^{-\frac{1}{2}-\alpha} \quad (2 \leq i \leq m \text{ and } m + n_{i-1} \leq k < m + n_i). \end{aligned}$$

For all other combinations of i and $k \geq 0$ we take $\hat{\phi}_i(k) = 0$, and we set $\hat{\phi}_i(-k) = \hat{\phi}_i(k)$. Here, $n_1 = 1$, $n_m = +\infty$, and the other n_i's are chosen such that for a fixed $R > 4$ (and all $m \geq 2$) we have

$$\|\phi_i\|_2^2 \leq RC^2 m^{-2\alpha-1}.$$

The reason that this is possible, is because there exists $R > 4$ such that

$$\sum_{k=m}^{\infty} C^2 k^{-1-2\alpha} \leq \frac{1}{4} RC^2 m^{-2\alpha} \quad (\forall\ m \geq 2).$$

It is clear that we can make sure that for $1 \leq i \leq m-1$,

$$\frac{1}{m} \sum_{k=m}^{\infty} C^2 k^{-1-2\alpha} - C^2 \cdot m^{-2\alpha-1} \leq \frac{1}{2} \|\phi_i\|_2^2 \leq \frac{1}{4} RC^2 m^{-2\alpha-1}.$$

This means that

$$\|\phi_m\|_2^2 \leq \frac{2}{m} \sum_{k=m}^{\infty} C^2 k^{-1-2\alpha} + 2C^2 \cdot m^{-2\alpha} \leq RC^2 m^{-2\alpha-1}.$$

Now define $f_m \in L^2[0,1]$ through its Fourier coefficients:

$$\hat{f}_m(k) = \begin{cases} 0 & \text{for } k = 0, \\ C|k|^{-\frac{1}{2}-\alpha} & \text{for } 1 \leq k \leq m-1 \text{ and } -m+1 \leq k \leq -1, \\ 0 & \text{for } |k| \geq m. \end{cases}$$

Define for $w \in \mathcal{W} = \{-1, 1\}^m$

$$\phi_w = \sum_{i=1}^{m} w_i \phi_i.$$

Clearly, $|\hat{f}_m(k) + \hat{\phi}_w(k)| = |\hat{g}(k)|$, so $f_m + \phi_w \in S_\alpha(K_0, M)$. Also, f_m and $\phi_1, \ldots, \phi_m$ are all orthogonal. Define

$$\mathbf{Q}_\theta^{\mathcal{W}} = 2^{-m} \sum_{w \in \mathcal{W}} \mathbf{P}_{\theta, f_m + \phi_w}.$$

According to (a rescaled version of) Lemma 3.2, if we wish to bound $\chi^2(\mathbf{Q}_\theta^{\mathcal{W}}, \mathbf{P}_{\theta, f_m})$, it is enough to bound

$$\sum_{i=1}^{m} \varepsilon^{-4} \|\phi_i\|_2^4 \leq R^2 C^4 m^{-1-4\alpha} \varepsilon^{-4}.$$

This remains bounded if we choose

$$m \approx \varepsilon^{-\frac{4}{1+4\alpha}}.$$

Now define

$$\mathbf{Q}_\tau^{\mathcal{W}} = 2^{-m} \sum_{w \in \mathcal{W}} \mathbf{P}_{\tau, f_m + \phi_w}.$$

Clearly, the above argument also shows that we have bounded $\chi^2(\mathbf{Q}_\tau^{\mathcal{W}}, \mathbf{P}_{\tau,f_m})$. Following the scheme depicted in Figure 1, we then need to bound $\chi^2(\mathbf{P}_{\theta,f_m}, \mathbf{P}_{\tau,f_m})$. This can be done by bounding

$$\varepsilon^{-2}\int_0^1 \left(f_m(x-\theta)-f_m(x-\tau)\right)^2 dx.$$

However,

$$\begin{aligned}
\varepsilon^{-2}\int_0^1 \left(f_m(x-\theta)-f_m(x-\tau)\right)^2 dx &= 2\varepsilon^{-2}\sum_{k=1}^{m-1}\left|e^{i2\pi k\theta}-e^{i2\pi k\tau}\right|^2 \hat{f}(k)^2 \\
&\leq 8\pi^2(\theta-\tau)^2\varepsilon^{-2}\sum_{k=1}^{m-1} k^2\hat{f}(k)^2 \\
&\lesssim (\theta-\tau)^2\varepsilon^{-2}m^{2-2\alpha} \lesssim (\theta-\tau)^2\varepsilon^{-\frac{10}{1+4\alpha}}.
\end{aligned}$$

This means that if we choose

$$\theta-\tau \approx \varepsilon^{\frac{5}{1+4\alpha}},$$

we can bound $\chi^2(\mathbf{P}_{\theta,f_m}, \mathbf{P}_{\tau,f_m})$, which proves the lower bound in Theorem 1.6 for the translation case, using the arguments belonging to Figure 1.

4 Appendix : technical lemmas

4.1 A general lemma

Let us state the following very useful lemma that we know from [6].

Lemma 4.1. *Let $Z(t)$ be a stochastic process differentiable a.s., μ and x positive real numbers and I an interval of $\mathbb{R}$. Then it holds*

$$\mathbf{P}\left(\sup_{\tau\in I} Z(\tau)>x\right) \leq \exp(-\mu x)\sup_{\tau\in I}\Big(\mathbf{E}\exp\{2\mu Z(\tau)\}\Big)^{1/2}\left(1+\mu\int_{\tau\in I}\{\mathbf{E}|Z'(\tau)|^2\}^{1/2}d\tau\right).$$

4.2 The center of symmetry case

Proof of Lemma 2.1 By definition,

$$\begin{aligned}
\eta_2(\tau)-K &= \sum_{k=1}^{K}\{\cos(2\pi k\tau)\xi_k+\sin(2\pi k\tau)\xi_k^*\}^2 \\
\eta_2'(\tau) &= 2\sum_{k=1}^{K}(2\pi k)\{\cos(2\pi k\tau)\xi_k+\sin(2\pi k\tau)\xi_k^*\}\{-\sin(2\pi k\tau)\xi_k+\cos(2\pi k\tau)\xi_k^*\}
\end{aligned}$$

To control their Laplace transforms note that for any fixed τ,

$$\eta_2(\tau) \stackrel{\mathcal{D}}{=} \sum_{k=1}^{K}\alpha_k^2 \quad\text{and}\quad \eta_2'(\tau)\stackrel{\mathcal{D}}{=} 2\sum_{k=1}^{K}(2\pi k)\alpha_k\tilde{\alpha_k} \stackrel{\mathcal{D}}{=} 2\sum_{k=1}^{K}(2\pi k)\left\{\beta_k^2-\tilde{\beta_k}^2\right\},$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution and $\alpha_k, \tilde{\alpha}_k, \beta_k, \tilde{\beta}_k$ are independent standard normal. Thus for any $0 < \mu < 1/8$, thanks to the fact that $-\log(1-v) \le v + v^2$ for $-1/2 < v < 1/2$,

$$\mathbf{E}(\exp\{2\mu(\eta_2 - K)\}) = \exp(-2\mu K)\exp\left\{-\sum_{k=1}^{K} \frac{1}{2}\log(1-4\mu)\right\} \le \exp(8\mu^2 K).$$

Similarly, there exists $C > 0$ such that

$$\mathbf{E}(\exp\{2\mu\eta_2'\}) \le \exp(C\mu^2 K^3).$$

By direct computation one checks $\mathbf{E}(\eta_2'(\tau)^2) \lesssim K^3$ and $\mathbf{E}(\eta_2''(\tau)^2) \lesssim K^5$. We apply Lemma 4.1 to the processes $\eta_2(\tau) - K$ and $\eta_2'(\tau)$ with the respective choices $x = K^{1/2}y$, $\mu = (K^{-1/2}y \wedge 1)/8$ and $x = K^{3/2}y$, $\mu = (K^{-3/2}y \wedge 1)/D$ (with D a large enough constant). $\square$

Proof of Lemma 2.2. We shall again apply Lemma 4.1. Let us denote

$$\zeta_1(\tau) \triangleq \frac{\eta_1(\tau) - \eta_1(\theta)}{v(\tau-\theta, K)^{1/2}}.$$

Note that due to (2.3), $\mathbf{E}(\zeta_1(\tau)^2) = 1$ thus $\mathbf{E}(\exp\{2\mu\zeta_1(\tau)\}) = \exp(2\mu^2)$. On the other hand, by direct calculation and using the inequality $(a+b)^2 \le 2a^2 + 2b^2$,

$$\begin{aligned}
\mathbf{E}(\zeta_1'(\tau)^2) &\le 2\frac{\mathbf{E}(\eta_1'(\tau)^2)}{v(\tau-\theta,K)} + 2\mathbf{E}(\{\eta_1(\tau)-\eta_1(\theta)\}^2)\left(\frac{\sum_{k=1}^{K}(2\pi k)f_k^2\sin(4\pi k\{\tau-\theta\})}{2v(\tau-\theta,K)^{3/2}}\right)^2 \\
&\lesssim K^3 v(\delta,K)^{-1} + \left(\sum_{k=1}^{K}(2\pi k)f_k^2\right)^2 v(\delta,K)^{-2} \lesssim K^3\gamma_\varepsilon^{-2} + K^2\|f\|^4\gamma_\varepsilon^{-4},
\end{aligned}$$

for any τ such that $\gamma_\varepsilon \le |\tau-\theta| \le 2\tau_0$, which concludes the proof of Lemma 2.2. $\square$

4.3 The period case, deterministic part

Before we start proving the Lemmas stated in Section 2.2, we formulate and prove a useful lemma about Fourier coefficients.

Lemma 4.2. *For any positive integer K, any $\tau, \theta \in \Theta^P$ and any $f = \{c_k\}_{k\in\mathbb{Z}}$ in $L^2[0,1]$,*

$$\begin{aligned}
\sum_{|k|\le K,\ l \ne]k\theta/\tau[} \frac{|c_{]k\theta/\tau[}||c_l|}{\left|l - \frac{k\theta}{\tau}\right|} &\le C\log K\left(1+\frac{\tau}{\theta}\right)^{1/2}\|f\|^2 \\
\sum_{|k|\le K}\left|\sum_{l\ne]k\theta/\tau[}\frac{|c_l|}{\left|\frac{l}{\theta}-\frac{k}{\tau}\right|}\right|^2 &\le C\log^2 K\|f\|^2,
\end{aligned}$$

where the summation over l is over all relative integers except $]k\theta/\tau[$.

Proof. Let us write

$$\begin{aligned}
\sum_{|k|\le K,\ l\neq]k\theta/\tau[} \frac{|c_{]k\theta/\tau[}||c_l|}{\left|l-\frac{k\theta}{\tau}\right|} &= \sum_{|k|\le K,\ .5\le|l-k\theta/\tau|\le K} \frac{|c_{]k\theta/\tau[}||c_l|}{\left|l-\frac{k\theta}{\tau}\right|} + \sum_{|k|\le K,\ |l-k\theta/\tau|>K} \frac{|c_{]k\theta/\tau[}||c_l|}{\left|l-\frac{k\theta}{\tau}\right|} \\
&= \mathcal{A}_1(K) + \mathcal{A}_2(K)
\end{aligned}$$

Using Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathcal{A}_2(K) &\le \sum_{|k|\le K} |c_{]k\theta/\tau[}| \left(\sum_{|l-k\theta/\tau|>K} |c_l|^2\right)^{1/2} \left(\sum_{|l-k\theta/\tau|>K} \frac{1}{\left|l-\frac{k\theta}{\tau}\right|^2}\right)^{1/2} \\
&\le \sum_{|k|\le K} |c_{]k\theta/\tau[}| \times \|f\| \times \left(\frac{C}{K}\right)^{1/2} \\
&\le \left(\sum_{|k|\le K} |c_{]k\theta/\tau[}|^2\right)^{1/2} \left(\sum_{|k|\le K} 1\right)^{1/2} \times \|f\| \times \left(\frac{C}{K}\right)^{1/2} \\
&\le C\left(1+\frac{\tau}{\theta}\right)^{1/2} \|f\|^2.
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_1(K) &\le \left(\sum_{|k|\le K,\ .5\le|l-k\theta/\tau|\le K} \frac{|c_{]k\theta/\tau[}|^2}{\left|l-\frac{k\theta}{\tau}\right|}\right)^{1/2} \left(\sum_{|k|\le K,\ .5\le|l-k\theta/\tau|\le K} \frac{|c_l|^2}{\left|l-\frac{k\theta}{\tau}\right|}\right)^{1/2} \\
&\le \left(\sum_{|k|\le K} |c_{]k\theta/\tau[}|^2 \sum_{.5\le|l-k\theta/\tau|\le K} \frac{1}{\left|l-\frac{k\theta}{\tau}\right|}\right)^{1/2} \left(\sum_{|k|\le K,\ .5\le|l-k\theta/\tau|\le K} \frac{|c_l|^2}{\left|l-\frac{k\theta}{\tau}\right|}\right)^{1/2} \\
&\le \left(\sum_{|k|\le K} |c_k|^2 \left(1+\frac{\tau}{\theta}\right) C\log K\right)^{1/2} \left(\sum_{|k|\le K} |c_k|^2 C\log\left\{\left(1+\frac{A}{B}\right)K\right\}\right)^{1/2}.
\end{aligned}$$

The second sum is bounded noting that the index l is always less in absolute value than $\left(1+\frac{A}{B}\right)K$, which proves the first part of the lemma. The second inequality follows from similar arguments separating the sum in l in two sums over $|l-k\theta/\tau|\le K$ and $|l-k\theta/\tau|>K$ respectively and is omitted. $\square$

Let us now turn to the proofs of Lemmas 2.3 and 2.4. Let $\hat\phi$ be the Fourier transform of the indicator function of $[-1/2,1/2]$: $\hat\phi(x)=\int_{-1/2}^{1/2} e^{2i\pi xt}dt = \sin(\pi x)/(\pi x)$. Note that there exists $C>0$ such that

$$|\hat\phi(u)|\le 1 \text{ for } u\in\mathbb{R}\ ,\quad |\hat\phi(u)|\le \frac{C}{|u|} \text{ for } |u|>1/4. \tag{4.1}$$

Let us introduce the auxiliary notation

$$a_{k,l}=\frac{T}{\theta}\left(l-\frac{k\theta}{\tau}\right)\quad ,\qquad b_{k,l}=\frac{T}{\theta}(l-k). \tag{4.2}$$

Proof of Lemma 2.3. Using the L^2-convergence of the Fourier series, one easily checks that for any f in $L^2[0,1]$ with Fourier coefficients $\{c_k\}$ and any $\tau \in \Theta^P$,

$$\int_{-T/2}^{T/2} \varepsilon_k(t/\tau) f(t/\theta) dt = T \sum_{l \in \mathbb{Z}} \hat{\phi}\left(\frac{T}{\theta}\left(\frac{k\theta}{\tau} - l\right)\right) \overline{c_l}.$$

Thus, with notation (4.2),

$$\Gamma(\tau) = T \sum_{|k| \leq K} \left| \sum_l c_l \hat{\phi}\left(a_{k,l}\right)\right|^2, \tag{4.3}$$

$$\Gamma(\tau) = T \sum_{|k| \leq K} \left[\left| c_{]k\theta/\tau[} \hat{\phi}\left(\frac{T}{\theta}\Delta(k\theta/\tau)\right) + \sum_{l \neq]k\theta/\tau[} c_l \hat{\phi}\left(a_{k,l}\right)\right|^2 \right]. \tag{4.4}$$

Let us expand the square in (4.4). The first term equals the main term in (2.12). The crossed term is upper bounded by, using (4.1),

$$\begin{aligned}
&2T \sum_{|k| \leq K} |c_{]k\theta/\tau[}| \left| \hat{\phi}\left(\frac{T}{\theta}\Delta(k\theta/\tau)\right)\right| \left| \sum_{l \neq]k\theta/\tau[} c_l \hat{\phi}\left(a_{k,l}\right)\right| \\
&\quad \leq 2 \sum_{|k| \leq K} |c_{]k\theta/\tau[}| \sum_{l \neq]k\theta/\tau[} |c_l| \frac{C}{\left|\frac{l}{\theta} - \frac{k}{\tau}\right|} \\
&\quad \leq 2C \sum_{|k| \leq K,\ l \neq]k\theta/\tau[} \frac{|c_{]k\theta/\tau[}||c_l|}{\left|\frac{l}{\theta} - \frac{k}{\tau}\right|}.
\end{aligned}$$

On the other hand the last term coming from the square in (4.4) is upper bounded by

$$T \sum_{|k| \leq K} \left| \sum_{l \neq]k\theta/\tau[} c_l \hat{\phi}\left(a_{k,l}\right)\right|^2 \leq \frac{1}{T} \sum_{|k| \leq K} \left| \sum_{l \neq]k\theta/\tau[} \frac{|c_l|}{\left|\frac{l}{\theta} - \frac{k}{\tau}\right|}\right|^2.$$

We conclude using Lemma 4.2. □

Proof of Lemma 2.4. We start the proof by noting that there exists a sequence Δ_T tending to $+\infty$ such that $\Delta_T \leq \inf(K, \gamma_T A/B)$ and $\gamma_T \Delta_T \varepsilon_T \to 0$, for instance we can take Δ_T equal to the minimum of $1/\sqrt{\gamma_T \varepsilon_T}$, K and $\gamma_T A/B$. We use this sequence to cut the sum defining $G(\tau)$ in two pieces.

$$\begin{aligned}
G(\tau) = \quad & T \sum_{|k| \leq \Delta_T} |c_{]k\theta/\tau[}|^2 \left| \hat{\phi}\left(\frac{T}{\theta}\Delta(k\theta/\tau)\right)\right|^2 &+ \quad & T \sum_{\Delta_T < |k| \leq K} |c_{]k\theta/\tau[}|^2 \left| \hat{\phi}\left(\frac{T}{\theta}\Delta(k\theta/\tau)\right)\right|^2 \\
= \quad & G_1(\tau) &+ \quad & G_2(\tau).
\end{aligned}$$

First let us note that

$$G_2(\tau) \leq T\left(1 + \frac{\tau}{\theta}\right) \sum_{\Delta_T < |k| \leq K} |c_k|^2.$$

This term is a $o(T)$ uniformly over $\mathcal{F}^+$, since $\sum_{|k|>\Delta_T} |c_k|^2 \le \sum_{|k|>\Delta_T} \{\frac{k}{\Delta_T}\}^{2\beta}|c_k|^2 \lesssim \Delta_T^{-2\beta}$.

Now let us study $G_1(\tau)$.

Assume that τ is not in a ball $B(j\theta/p, \varepsilon_T)$ with $p \le \gamma_T$ and $j \ge 1$. Then for any integer p such that $1 \le p \le \gamma_T$,

$$\left|\frac{k\theta}{\tau} - p\right| > p\varepsilon_T.$$

Since in the sum defining G_1, $|k| \le \Delta_T$,

$$]k\theta/\tau[\le \frac{kB}{A} \le \frac{\Delta_T B}{A}.$$

If $]k\theta/\tau[> 0$ then since $\Delta_T B \le A\gamma_T$, the preceding lines imply that

$$\left|\frac{T}{\theta}\Delta(k\theta/\tau)\right| > \frac{T\varepsilon_T}{\theta}.$$

If $]k\theta/\tau[= 0$, then note that

$$\left|\frac{T}{\theta}\Delta(k\theta/\tau)\right| \ge |k|T/\tau.$$

In all cases, since $T\varepsilon_T \to +\infty$,

$$|G_1(\tau) - |c_0|^2 T| \le T\left(1 + \frac{\tau}{\theta}\right)\|f\|^2 \left(\frac{\theta^2}{T^2\varepsilon_T^2} + \frac{1}{T^2}\right).$$

Assume that τ is in a ball $B(j\theta/p, \varepsilon_T)$ with $p \le \gamma_T$ and $gcd(p, j) = 1$. Then there are two cases for the integer k in the sum defining $G_1(\tau)$.

- There exists an integer q such that $k = qj$. Then

  $$|k\theta/\tau - pq| < pq\varepsilon_T \le \gamma_T\Delta_T\varepsilon_T/j \le \gamma_T\Delta_T\varepsilon_T.$$

  Since $\gamma_T\Delta_T\varepsilon_T = o(1)$ as $T \to +\infty$ it holds for T large enough that $]k\theta/\tau[= pq$.

- The integer k is not a multiple of j. Then $k = qj + r$ with $r < j$. Note that

  $$\left|\frac{k\theta}{\tau} - \left(pq + \frac{rp}{j}\right)\right| \le pq\varepsilon_T + \frac{rp\varepsilon_T}{j} \le Cpq\varepsilon_T \le C\frac{\gamma_T\Delta_T\varepsilon_T}{j}.$$

  Since $gcd(p, j) = 1$, $k\theta/\tau$ is at a distance to the integers larger than $1/2j$, that is

  $$\Delta(k\theta/\tau) \ge \frac{1}{2j}.$$

  Since τ must lie in Θ^P, we have that $j\theta/p \le 2B$. Thus $j^{-1} \ge \theta/(2pB)$. Therefore

  $$\left|\frac{T}{\theta}\Delta(k\theta/\tau)\right| \ge \frac{T}{2j\theta} \ge \frac{T}{4\gamma_T B}.$$

  We have assumed that $\gamma_T T^{-1} \to 0$ as $T \to +\infty$, thus

  $$\left|\hat{\phi}\left(\frac{T}{\theta}\Delta(k\theta/\tau)\right)\right|^2 \le C\left(\frac{\gamma_T B}{T}\right)^2.$$

Finally we obtain that if τ is in a ball $B(j\theta/p, \varepsilon_T)$,

$$G_1(\tau) \leq T \sum_{|q|\leq \Delta_T} |c_{pq}|^2 + T\left(\frac{\gamma_T B}{T}\right)^2 \left(1+\frac{\tau}{\theta}\right) C\|f\|^2.$$

Equation (2.16) follows from the definition of the function G. $\square$

4.4 Processes in the period estimation case

The process ζ is Gaussian and its covariance structure is very much related to the deterministic part Γ as is established in the next lemma.

Lemma 4.3. *For any $\tau \in \Theta^P$, as T tends to $+\infty$,*

$$\mathbf{E}(\zeta(\tau)^2) = 4\Gamma(\tau) + \tau\|f\|^2\mathcal{O}(\log K). \tag{4.5}$$

The process Ψ is apparently more complicated than its counterpart η_2 in the translation case but it can still by written as (infinite) weighted sum of squares of independent standard Gaussian variables. For any real s and any $\tau \in \Theta^P$, let us define the kernel $K_T^\tau(s)$ as the function $s \to T^{-1}\sum_{|k|\leq K} \varepsilon_k\{s/\tau\}$.

Lemma 4.4. *For any $\tau \in \Theta^P = [A, B]$, there exist a sequence of real numbers $\{\beta_l(\tau)\}_{l\geq 1}$ and a sequence of independent standard normal random variables $\alpha_l(\tau)$ such that*

$$\Psi(\tau) = \sum_{l=1}^{+\infty} \beta_l(\tau)\, \alpha_l^2(\tau). \tag{4.6}$$

The $\beta_k(\tau)'s$ are the eigenvalues of the operator on $L^2([-T/2, T/2])$ defined by

$$g \to \left\{ t \to \int_{-T/2}^{T/2} K_T^\tau(t-u)g(u)du \right\}.$$

Similar results hold for the derivatives $\Psi^{(p)}$ of the process Ψ, with corresponding eigenvalues $\beta_l^{(p)}(\tau)$. If **Var** *denotes the variance, it holds*

$$\begin{aligned}
\mathbf{E}[\Psi^{(p)}(\tau)] &= \sum_{l=1}^{+\infty} \beta_l^{(p)}(\tau) = \int_{-T/2}^{T/2} K_T^{\tau\ (p)}(0)dt && (4.7)\\
\mathbf{Var}[\Psi^{(p)}(\tau)] &= 2\sum_{l=1}^{+\infty} \beta_l^{(p)}(\tau)^2 = 2\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \left|K_T^{\tau\ (p)}(t-u)\right|^2 dtdu, && (4.8)
\end{aligned}$$

The following lemma gives precise control on the preceding quantities for $p = 0, 1$.

Lemma 4.5. *The process Ψ has mean $2K$, the process Ψ' is centered. Moreover, as T tends to $+\infty$, for any $\tau \in \Theta^P$,*

$$\begin{aligned}
\mathbf{E}(\{\Psi(\tau) - 2K\}^2) &= 4K + \mathcal{O}\left(K\frac{\tau^2}{T^2}\right). && (4.9)\\
\mathbf{E}(\Psi'(\tau)^2) &= T^2\tau^{-4}\{1+o(1)\} \sum_{|k|\leq K} (2\pi k)^2/3. && (4.10)
\end{aligned}$$

Proof of Lemma 4.3 Let us denote $\gamma_k = \sum_l c_l \hat{\phi}(a_{k,l})$, with $a_{k,l} = \frac{T}{\theta}\left(l - \frac{k\theta}{\tau}\right)$.

$$\begin{aligned}
\mathbf{E}(\zeta(\tau)^2) &= 4T \int_{-T/2}^{T/2} \left| \sum_{|k| \le K} \left\{ \sum_l c_l \hat{\phi}(a_{k,l}) \right\} \varepsilon_k(tT/\tau) \right|^2 dt \\
&= 4T \left\{ \sum_{|k| \le K} |\gamma_k|^2 + \sum_{|k|,|p| \le K, k \ne p} \gamma_k \overline{\gamma_p} \hat{\phi}\left(\frac{(k-p)T}{\tau}\right) \right\}
\end{aligned}$$

In the first term of the sum we recognize $4\Gamma(\tau)$ (see Lemma 2.3). To see that the second term is negligible use Cauchy-Schwarz inequality as follows

$$T \sum_{|k|,|p| \le K, k \ne p} \gamma_k \overline{\gamma_p} \hat{\phi}\left(\frac{(k-p)T}{\tau}\right) \lesssim \left(\sum_{|k|,|p| \le K, k \ne p} \frac{|\gamma_k|^2}{|k-p|} \right) \lesssim \log K \frac{\Gamma(\tau)}{T}. \quad \square$$

Proof of Lemma 4.4. The statement follows from the fact that the considered operator is selfadjoint and compact, see [2] for details.

Proof of Lemma 4.5. From the definition of Ψ it follows

$$\begin{aligned}
\mathbf{Var}(\Psi(\tau)) &= \frac{2}{T^2} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \left| \sum_{|k| \le K} \varepsilon_k\{(t-u)/\tau\} \right|^2 dtdu \\
&= 2 \sum_{|k|,|p| \le K} \int_{-1/2}^{1/2} \varepsilon_{k-p}(Tt/\tau)dt \int_{-1/2}^{1/2} \varepsilon_{p-k}(Tu/\tau)du \\
&= 2 \sum_{|k|,|p| \le K} \left(\mathbf{1}_{k=p} + \hat{\phi}\left(\frac{(k-p)T}{\tau}\right)^2 \mathbf{1_{k \ne p}} \right) = 4K + \mathcal{O}\left(\frac{\tau^2}{T^2} \sum_{|k|,|p| \le K, k \ne p} \frac{1}{(k-p)^2} \right).
\end{aligned}$$

Thanks to (4.8),

$$\mathbf{E}(\Psi'(\tau)^2) = 8\pi^2 T^2 \tau^{-4} \sum_{|k|,|p| \le K} kp \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \exp(2i\pi(t-u)(k-p)/\tau)(t-u)^2 dtdu$$

The term $k = p$ in the preceding sum equals $T^2 \tau^{-4} \sum_{|k| \le K} (2\pi k)^2/3$. It remains to see that the term $k \ne p$ is negligible.

$$\begin{aligned}
&\sum_{k \ne p} kp \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \exp(2i\pi(t-u)(k-p)/\tau)(t-u)^2 dtdu \\
&= \frac{\tau^2}{T^2} \sum_{k \ne p} kp \Bigg\{ \int_{-1/2}^{1/2} t^2 \varepsilon_{k-p}(tT/\tau)dt \int_{-1/2}^{1/2} \varepsilon_{k-p}(-uT/\tau)du \\
&\quad -2 \int_{-1/2}^{1/2} t\varepsilon_{k-p}(tT/\tau)dt \int_{-1/2}^{1/2} u\varepsilon_{k-p}(-uT/\tau)du \\
&\quad + \int_{-1/2}^{1/2} \varepsilon_{k-p}(tT/\tau)dt \int_{-1/2}^{1/2} u^2 \varepsilon_{k-p}(-uT/\tau)du \Bigg\} = \mathcal{O}(\tau^2/T^2) \sum_{k \ne p} \left| \frac{kp}{(k-p)^2} \right|.
\end{aligned}$$

This yields the result noting that the double sum in the preceding display is a $\mathcal{O}(K^3)$. $\square$

Proof of Lemma 2.5 With the tools provided by Lemmas 4.3,4.4 and 4.5, together with the application of Lemma 4.1, the proof is similar to the ones of Lemmas 2.1 and 2.2 and is omitted. $\square$

Proof of Lemmas 2.7 and 2.8 Since $K < \lambda^{-1}$, $]k\theta/\tau[= k$ and Lemma 2.3 yields

$$\Gamma(\tau) - \Gamma(\theta) = T \sum_{|k|\le K} |c_k|^2 \left(\left|\hat{\phi}\left(k\lambda T\right)\right|^2 - 1 \right) + \mathcal{O}(l_T).$$

The first lemma is proved noting that

$$(1 - ck^2\lambda^2T^2)\mathbf{1}_{|k|<\lambda^{-1}T^{-1}} \le \left|\hat{\phi}(k\lambda T)\right|^2 \le (1 - Ck^2\lambda^2T^2)\mathbf{1}_{|k|<\lambda^{-1}T^{-1}} + \frac{1}{2}\mathbf{1}_{\lambda^{-1}T^{-1}\le|k|<K}.$$

To prove Lemma 2.8, with the notation $\langle\varepsilon_k(\cdot/\tau), W'\rangle \triangleq \int_{-T/2}^{T/2}\varepsilon_k(\cdot/\tau)dW(t)$,

$$\begin{aligned}
\zeta(\tau) - \zeta(\theta) &= 2\sum_{|k|\le K} c_k \left[\hat{\phi}\left(\frac{T}{\theta}k\left(\frac{\theta}{\tau} - 1\right)\right) - 1\right]\langle\varepsilon_k(t/\tau), W'\rangle \\
&\quad +2\sum_{|k|\le K} c_k\langle\varepsilon_k(t/\tau) - \varepsilon_k(t/\theta), W'\rangle \\
&\quad +2\sum_{|k|\le K}\sum_{l\ne k} c_l\left[\hat{\phi}\left(\frac{T}{\theta}\left\{\frac{k\theta}{\tau} - l\right\}\right)\langle\varepsilon_k(t/\tau), W'\rangle - \hat{\phi}\left(\frac{T}{\theta}\left\{k - l\right\}\right)\langle\varepsilon_k(t/\theta), W'\rangle\right] \\
&= \zeta_1(\tau) + \zeta_2(\tau) + \zeta_3(\tau).
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbf{E}(\zeta_1(\tau)^2) &= 4T\sum_{|k|\le K}|c_k|^2\left\{1 - \hat{\phi}(k\lambda T)\right\}^2 \\
&\quad + \sum_{-K\le p\ne q\le K} c_pc_q\{\hat{\phi}(p\lambda T) - 1\}\{\hat{\phi}(q\lambda T) - 1\}T\hat{\phi}\left(\frac{T}{\tau}(p-q)\right).
\end{aligned}$$

The second term can be bounded by the sum of $|c_pc_q|/|p-q|$ and thus also by $\mathcal{O}(\log K)$. This implies that

$$c\mathcal{V}(\lambda, K) + \mathcal{O}(\log K) \le \mathbf{E}(\zeta_1(\tau)^2) \le C\mathcal{V}(\lambda, K) + \mathcal{O}(\log K).$$

We also have

$$\begin{aligned}
\mathbf{E}(\zeta_2(\tau)^2) &= 4\sum_{|k|\le K}|c_k|^2\langle\varepsilon_k(t/\tau) - \varepsilon_k(t/\theta), \varepsilon_k(t/\tau) - \varepsilon_k(t/\theta)\rangle \\
&\quad +4\sum_{-K\le p\ne q\le K} c_pc_q\langle\varepsilon_p(t/\tau) - \varepsilon_p(t/\theta), \varepsilon_q(t/\tau) - \varepsilon_q(t/\theta)\rangle.
\end{aligned}$$

The first term is upper bounded by $CT\sum_{|k|\le K}|c_k|^2\left\{1 - \hat{\phi}(k\lambda T)\right\}$. The second term is upper bounded by $\mathcal{O}(\log K)$ as above.

Finally, the terms appearing in the variance of ζ_3 are of the form

$$\sum_{-K\leq k,p\leq K}\sum_{l,q} c_l c_q \hat{\phi}\left(\frac{T}{\theta}\left(\frac{k\theta}{\tau}-l\right)\right) T\hat{\phi}\left(\frac{T}{\theta}\left(\frac{k\theta}{\tau}-p\right)\right)\hat{\phi}\left(\frac{T}{\theta}(p-q)\right)$$

and thus are upper bounded in modulus by

$$\frac{1}{T^2}\sum_{-K\leq k,p\leq K}\frac{1}{|k-p|}\sum_{l,q}\frac{|c_l||c_q|}{|k-l||p-q|}\leq\frac{K\log K}{T^2}.$$

Since $K=o(T^2)$, (2.21) is established. Finally, to prove (2.22), use Lemmas 4.4 and 4.1 as in the proof of Lemma 2.1. □

Acknowledgments. The authors would like to thank the Statistics department of the Université Paris-Sud for offering them the possibility to work together on this subject. The first author would like to thank Professor Aad van der Vaart for stimulating discussions.

References

[1] Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner. *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1993.

[2] I. Castillo. Penalized profile likelihood methods and second order properties in semiparametrics. *PhD Thesis, Université Paris-Sud*, 2006.

[3] I. Castillo. Semi-parametric second-order efficient estimation of the period of a signal. *Bernoulli*, 13(4):910–932, 2007.

[4] A. Dalalyan, G. Golubev, and A. Tsybakov. Penalized maximum likelihood and semiparametric second order efficiency. *Ann. Statist.*, 34(1):169–201, 2006.

[5] A. S. Dalalyan. Stein shrinkage and second-order efficiency for semiparametric estimation of the shift. *Math. Methods Statist.*, 16(1):42–62, 2007.

[6] G. K. Golubev. Estimating the period of a signal of unknown shape corrupted by white noise. *Problems of Information Transmission*, 24(4):288–299, 1988.

[7] I. A. Ibragimov and R. Z. Has'minskii. *Statistical estimation: asymptotic theory.* Springer, New York e.a., 1981.

[8] Marc Lavielle and Céline Lévy-Leduc. Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals. *IEEE Trans. Signal Process.*, 53(7):2306–2314, 2005.

[9] Catherine Matias. Semiparametric deconvolution with unknown noise variance. *ESAIM Probab. Statist.*, 6:271–292 (electronic), 2002. New directions in time series analysis (Luminy, 2001).

[10] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1):28–76, 1994.

[11] A. W. van der Vaart. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. Cambridge, 1998.

[12] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[13] Grace Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.