Estimation of the distribution of random shifts deformation

I. Castillo & J-M. Loubes

Abstract

The observations are discrete values of functions shifted by translation effects. These unobserved translation parameters are i.i.d. realizations of a random variable with unknown distribution, modelling the variability in the response of each individual. Our aim is to construct a nonparametric estimator of the density of these random translation deformations using semiparametric preliminary estimates of the shifts. First, a theoretical study of the semiparametric procedure is carried out. Second order results are obtained and a practical estimate is built based on second order considerations. Second, we use the semiparametric estimators to construct the nonparametric estimator of the density. Both rate of convergence and an algorithm to construct the estimator are provided.

Keywords: Semiparametric statistics, Order two properties, Penalized Maximum Likelihood, Practical algorithms .
Subject Class. MSC-2000: 62G05, 62G20 .

1 Introduction

Our aim is to estimate the density φ of random variables θ_j, $j = 1, \ldots, J_n$, observed in a panel data analysis framework, in particular cases of the following general model. Consider J_n unknown curves $t \to f^{[j]}(\theta_j, t)$, indexed by i.i.d. random parameters θ_j, observed at multiple points t_{ij}, $i = 1, \ldots, n$ in the following regression framework

$$Y_{ij} = f^{[j]}(\theta_j, t_{ij}) + \sigma\varepsilon_{ij} \ , \ i = 1, \ldots, n, \ j = 1, \ldots, J_n, \quad (1)$$

where ε_{ij} are i.i.d. random noise. The number of points per curve is denoted by n while J_n stands for the number of curves (or subjects). Assumptions on the design and the form of the curves are made precise in the sequel. Equation (1) describes the situation where the outcome of an experiment depends on a random variable θ which models the fact that each subject j can react in a different way within a mean behaviour, with slight variations given by the unknown curves $f^{[j]}$. Provided identifiability of the parameters holds, estimating φ the density of the unobserved θ_j's enables to understand this mean behaviour.

A large number of practical cases in biology, data mining or econometrics may be stated in the form of estimating the density of unobserved random variables defined as in model (1). It is in particular the case when data variations take into account the variability of the individual.

In functional data mining for instance, data clustering leads to different homogeneous groups, each representing a specific mass behavior, but curves within one group still differ slightly the one from another. More precisely, there is a mean unknown pattern, such that each observation curve is warped from this archetype by a random parameter θ_j. Estimation of these parameters have been investigated by several authors in a very general framework. Among this work, we refer to nonparametric methods in [12, 14], or Dynamic Time Warping in [24] and [19]. However little attention is paid to the law of these random parameters which is needed to fully understand the clustering procedure. Moreover sharp estimates of such parameters are required to achieve this nonparametric density estimation.

In biology, Equation (1) models the concentration of a medicine in blood, often referred as the "pharmacokinetics". The functions $f^{[j]}$ usually arises from mechanistic structural models and describes the evolution of the concentration of the studied medicine. The θ parameter makes the kinetic individual dependent, which corresponds to the biological specificity of each individual, with respect to the treatment. Hence, estimating the law of θ is of crucial importance to study the effects of individuals to a particular medicine. We refer to [10], [18], [20] or [7] for more references. Unfortunately, the knowledge of the functions $f^{[j]}$, or a parametric model for the density φ are needed.

Finally, the problem is twofold: first, provided the model is identifiable, estimate the unknown realizations θ_j warped by the unknown functions $f^{[j]}$; then, use these estimates to build a nonparametric estimator of their density. So, nonparametric estimation of φ belongs to the class of inverse problems for which the subject of the inversion is a probability measure. Since the $f^{[j]}$'s are unknown, the underlying inverse problem becomes more than harmful as sharp approximations of the θ_j are needed, preventing flawed rates of convergence for the density estimator. While the estimation of parameters, observed through their image by an operator, traditionally relies on the inversion of the operator, here the repetition of the observations enables to use recent advances in semiparametric estimation to improve the usual strategies developed to solve such a problem.

In the very general framework (1), the problem seems very difficult to solve unless a special dependency between $f^{[j]}$ and θ_j is specified. In this article we study in detail the case where the warping effect of the parameter θ_j is a translation effect, that is $f^{[j]}(\theta_j, t) = f^{[j]}(t - \theta_j)$. For θ_j, $j = 1, \ldots, J_n$ i.i.d. parameters with distribution μ and density with respect to Lebesgue measure φ, consider the observations

$$Y_{ij} = f^{[j]}(t_{ij} - \theta_j) + \sigma\varepsilon_{ij} \qquad i = 1, \ldots, n\ , j = 1, \ldots, J_n\ , \tag{2}$$

where $f^{[j]}$ are *symmetric* functions satisfying some additional assumptions. Our approach consists, first, in the estimation of the shifts θ_j while the functions $f^{[j]}$ play the role of nuisance parameters. We follow the semiparametric approach introduced in [8] in the Gaussian white noise framework and extend it to the regression framework. This provides sharp estimators of the unobserved shifts, up to order 2 expansions. Alternative methods can be found in [11] or [23]. These preliminar estimates enable, in a second time, to recover the unknown density φ of the θ_j's as if the shifts were directly observed, at least if J_n is not significatively larger than n. This paper also provides a practical algorithm, for both the semiparametric and the nonparametric steps. The first step is the most difficult one: to build practicable semiparametric estimators, we propose an algorithm which refines the one proposed in [16] for the period model and relies on the previously obtained second order expansion.

Beyond the shift estimation case, which involves a symmetry assumption on $f^{[j]}$, our procedure may be applied to semiparametric models where an explicit penalized profile likelihood is available and well-behaved estimators of the θ_j's can be obtained. A particularly important example in applications is the estimation of the period of an unknown periodic function (see [16]). Given a sequence of J_n experiments like the one considered in [16], one might be interested in estimating the law of the corresponding periods of the signals. Then our method also applies, under some restrictions made explicit in the sequel.

The paper falls into the following parts. In Section 2, semiparametric estimators $\hat{\theta}_j$ of the realizations of the shift parameter in model (2) are proposed and sharp bounds between $\hat{\theta}_j$ and θ_j are provided. Then, in Section 3, a nonparametric estimator of the unknown distribution is considered while rates of convergence are provided in the case where μ admits a density, in the general model (1) under the condition that the θ_j's can be sufficiently well approximated. In Section 4, the practical estimation problem is considered and a simulation study is conducted in model (2). Technical proofs are gathered in Section 5.

2 Semiparametric Estimation of the shifts

In this section, we provide, for each fixed j, semiparametric estimators of the j^{th} realization θ_j of the random variable θ, observed in Model (2). Hence, conditionally to the event $\theta_j = \theta$, we construct an estimator $\widehat{\theta}_j$ and establish asymptotic results for the conditional distribution $\left(\widehat{\theta}_j \mid \theta_j = \theta\right)$, gathered in Lemmas 1, 2 and 3. More precisely, in the discrete-time model at hand, we obtain second-order semiparametric expansions for a family of estimators indexed by a sequence of weights.

In the remaining of this section, since j is fixed, the index j in the notation is dropped (for instance Y_{ij} is simply denoted by Y_i). We shall denote by $\|.\|$ the L^2-norm on $[0,1]$ and by $\|.\|_\infty$ the L^∞-norm on $\mathbb{R}$.

2.1 Shift estimation in the discrete time translation model

The model reduces to

$$Y_i = f(t_i - \theta) + \varepsilon_i \qquad i = 1, \dots, n, \tag{3}$$

where f is a *symmetric* function satisfying some additional assumptions detailed in the next subsection. For each fixed j, the corresponding problem is the one of semiparametric estimation of the center of symmetry in a discrete framework.

Working assumptions in the translation model

We assume that the support Θ of the distribution μ of the random variable θ is compactly supported, the support being contained in an interval of diameter upper-bounded by $1/2$

$$(A1) \quad \Theta = \{\theta,\ |\theta| \leqslant \tau_0\}, \quad \text{where } \tau_0 \ \text{ is such that } 0 < \tau_0 < 1/4.$$

The function f is assumed to be symmetric and periodic with period 1 with Fourier coefficients denoted by f_k, $k \geqslant 1$,

$$(A2) \quad f(t) = \sqrt{2} \sum_{k \geqslant 1} f_k \cos(2\pi k t), \quad \text{where} \quad f_k = \sqrt{2} \int_0^1 f(t) \cos(2\pi k t) dt.$$

We assume that there exist $\rho > 0$ and $C_0 < +\infty$ such that f belongs to the set F defined by

$$(A3) \quad F = F(\rho, C_0) = \{f \in \mathcal{C}^2(\mathbb{R}), \quad f_1^2 \geqslant \rho, \quad \|f''\|^2 \leqslant C_0\}.$$

Conditions (A2) and (A3) can be seen as working assumptions. Assuming periodicity of f is not a drawback since, in practice, the function f is compactly supported and can easily be periodicized. The assumption that f is $\mathcal{C}^2$ is handy in particular for proving the second order properties of the estimator. Note that we assume nothing more than in [8].

Identifiability in model (2) follows from : symmetry, 1-periodicity of the functions (note that assuming that $f_1^2 \geqslant \rho$ implies that f cannot be periodic of smaller period) and the fact that the diameter of Θ is less than 1/2.

Note also that within this framework, the Fisher information for estimating θ_j for a fixed j is, as n tends to $+\infty$, given by $\{1 + o(1)\} n \|f'\|^2$.

Construction of the estimator

Let us introduce some auxiliary notation

$$x_k = \frac{1}{n} \sum_{i=1}^{n} \sqrt{2} \cos(2\pi k t_i) Y_i, \qquad \xi_k = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^{n} \sqrt{2} \cos(2\pi k t_i) \varepsilon_i.$$

$$x_k^* = \frac{1}{n} \sum_{i=1}^{n} \sqrt{2} \sin(2\pi k t_i) Y_i, \qquad \xi_k^* = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^{n} \sqrt{2} \sin(2\pi k t_i) \varepsilon_i.$$

Note that the sequences $\{x_k\}$ and $\{x_k^*\}$ are observed. This leads to

$$x_k = \cos(2\pi k \theta) f_k + d_{k,n} + \frac{1}{\sqrt{n}} \xi_k, \tag{4}$$

$$x_k^* = \sin(2\pi k \theta) f_k + d_{k,n}^* + \frac{1}{\sqrt{n}} \xi_k^*, \tag{5}$$

and $d_{k,n}$ and $d_{k,n}^*$ are terms of difference between the Fourier coefficient and its approximation

$$d_{k,n} = \sqrt{2} \left(\frac{1}{n} \sum_{i=1}^{n} \cos(2\pi k t_i) f(t_i - \theta) - \int_0^1 \cos(2\pi k t) f(t - \theta) dt \right).$$

The term $d_{k,n}^*$ is obtained in a similar way replacing the cosine by a sine. Both are bounded uniformly in θ by $\|f'\|_\infty / n$ (thus the dependency of $d_{k,n}$ in θ is omitted). The variables $(\xi_k, \xi_k^*)_{k \geqslant 1}$ are Gaussian random variables, possibly correlated.

To simplify the calculations, *assume that* $t_i = i/n$ so that, using the orthogonality of the trigonometric basis over this system of points, the preceding variables are independent standard Normal. However, handling the case of non equally spaced observation times does not change the flavour of the proof under some regularity assumptions for the function f.

To build the estimator of θ, a penalized profile likelihood method is used, as in [8]. Let us recall the idea of the construction. For each integer k, define a penalized likelihood $p_\theta(x_k, x_k^*, f_k)$ as

$$\left(\frac{1}{\sqrt{2\pi}}\right)^3 \exp\left(-\frac{n}{2}(x_k - \cos(2\pi k\theta)f_k - d_{k,n})^2 - \frac{n}{2}(x_k^* - \sin(2\pi k\theta)f_k - d_{k,n}^*)^2 - \frac{f_k^2}{2\sigma_k^2}\right).$$

This is the usual likelihood corresponding to the observation (x_k, x_k^*) with an additional penalization term $-f_k^2/2\sigma_k^2$, where σ_k has to be chosen. Then, as in the profile likelihood technique (see [22, Chap. 25]), we "profile out" the nuisance parameter f_k by setting

$$f_k^*(\theta) = \underset{f_k}{\operatorname{argmax}}\ p_\theta(x_k, x_k^*, f_k)$$
$$\widehat{\theta}_{PML} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ p_\theta(x_k, x_k^*, f_k^*(\theta)).$$

The problem here is that $d_{k,n}$ depends on f_k but, as it is a $O(n^{-1})$ uniformly in θ, we do not consider its dependance in f_k and hence, with this assumption

$$f_k^*(\theta) \approx \frac{\sigma_k^2}{\sigma_k^2 + n^{-1}} \left(x_k \cos(2\pi k\theta) + x_k^* \sin(2\pi k\theta) - d_{k,n}\cos(2\pi k\theta) + d_{k,n}^* \sin(2\pi k\theta)\right)$$
$$\widehat{\theta}_{PML} \approx \underset{\tau\in\Theta}{\operatorname{argmax}} \sum_{k\geqslant 1} \frac{\sigma_k^2}{\sigma_k^2 + n^{-1}} \left(\frac{1}{n}\sum_{i=1}^n \cos(2\pi k(t_i - \tau))Y_i + r_{k,n}(\tau)\right)^2,$$

where

$$r_{k,n}(\tau) = \cos(2\pi k\tau)d_{k,n} + \sin(2\pi k\tau)d_{k,n}^* = \frac{1}{n}\sum_{i=1}^n \cos(2\pi k(t_i-\tau))f(t_i-\tau) - f_k.$$

This is not yet an estimator since we do not know $r_{k,n}$, but this quantity is a $O(n^{-1})$ uniformly in θ, hence we can expect that it will not be crucial when computing the Argmax (as shown in Section 2.2, forgetting this term does not matter for consistency with a sufficiently good rate). Now note that $\sigma_k^2/(\sigma_k^2 + n^{-1})$ describes $[0,1]$ (the value 1 corresponds to the choice $\sigma_k = +\infty$). Our definitive estimator is then

$$\widehat{\theta} = \underset{\tau\in\Theta}{\operatorname{argmax}} \sum_{k\geqslant 1} h_k \left(\frac{1}{n}\sum_{i=1}^n \cos(2\pi k(t_i - \tau))Y_i\right)^2, \tag{6}$$

where (h_k) is a sequence of real numbers in $[0,1]$ satisfying some conditions made precise in the following subsection. The sequence (h_k) is called *sequence of weights* or *filter*. The estimator $\widehat{\theta}$ has to be compared to the estimator $\widehat{\theta}_{PML}$ in [8]: here the integral in the definition of $\widehat{\theta}_{PML}$ is replaced by the equivalent discrete sum in the discrete-time model. Thus we will have to handle with approximations of the Fourier coefficients of the function f instead of the $f_k's$.

2.2 Asymptotic behavior of the shifts estimators

The behavior of the estimates $\widehat{\theta}_j$ depends on proper choices of the sequence of weights.

Assumptions on the sequence of weights (h_k)

The sequence (h_k) satisfies: $h_1 = 1$, $0 \leqslant h_k \leqslant 1$ for all $k \geqslant 1$, and there are positive constants D_1 and ρ_1 such that:

$(C1)$ The number of weights such that $h_k \neq 0$ is finite.

$$(C2) \quad \left[\sum_{k\geqslant 1}(2\pi k)^2 h_k^2\right]^{1/2} \geqslant \rho_1(\log^2 n)\max_{k\geqslant 1}(2\pi k)h_k.$$

$$(C3) \quad \sum_{k\geqslant 1} h_k(2\pi k)^4 \leqslant D_1 n.$$

$$(T) \quad \left(\sum_{k\geqslant 1}(1-h_k)(2\pi k)^2 f_k^2\right)^2 = o\left(\sum_{k\geqslant 1}(1-h_k)^2(2\pi k)^2 f_k^2\right).$$

The first condition is quite natural to make the estimator feasible, Conditions (C2) and (C3) precise the range of the sequence (h_k). Condition (T) allows to obtain second order properties (see the proof of Lemma 2).

Remark 1. As noted in [8], conditions (C1), (C2), (C3) and (T) are fulfilled for a quite wide range of weights. For instance, the sequences $(h_k = \mathbf{1}_{1\leqslant k\leqslant N(T)})$, also called projection weights, satisfy the preceding conditions since (C2) and (C3) are satisfied respectively for $N(T) \geqslant C\log^4 n$ and $N(T) \leqslant Cn^{1/5}$, while condition (T) is always satisfied for projection weights since $\sum_{k\geqslant N(T)}(2\pi k)^2 f_k^2 \to 0$, as $n \to +\infty$, thanks to (A3).

Asymptotic properties

For easyness of reference in Section 3, it is convenient here to make the dependance in j explicit again. Then in the following lemmas $f^{[j]\prime}$ and $f_k^{[j]}$ respectively denote the derivative of $f^{[j]}$ and the Fourier coefficients of $f^{[j]}$.

Lemma 1 (Deviation bound). *Assume that* (A), (C), (T) *are fulfilled. For any* $K > 0$ *and any positive integer* n, *denote* $x_n = K\sqrt{\log n}$. *There exist positive constants* c_1, c_2 *such that for any* $K > 0$, *for* n *large enough, uniformly in* $j \in \{1, \dots, J_n\}$, $\theta_j \in \Theta$ *and* $f^{[j]} \in F$, *it holds*

$$\mathbf{P}\left(\sqrt{n}|\widehat{\theta}_j - \theta_j| > x_n|\theta_j\right) \leqslant c_1\exp(-c_2 x_n^2). \tag{7}$$

Lemma 2 (Second order Expansion). *Assume* (A), (C), (T) *and set* $R^n[h, f^{[j]}] = \sum_{k=1}^{\infty}(2\pi k)^2[(1-h_k)^2 f_k^{[j]\,2} + h_k^2/n]$. *Uniformly in* $j \in \{1, \dots, J_n\}$, $\theta_j \in \Theta$ *and* $f^{[j]} \in F$, *as* n *tends to* $+\infty$,

$$\mathbf{E}\left((\widehat{\theta}_j - \theta_j)^2|\theta_j\right) = \frac{1}{n\|f^{[j]\prime}\|^2}\left(1 + (1+o(1))\frac{R^n[h, f^{[j]}]}{\|f^{[j]\prime}\|^2}\right). \tag{8}$$

Lemma 3 (Asymptotical Bias). *Assume* (A), (C), (T), *then, uniformly in* $j \in \{1, \dots, J_n\}$, *as* n *tends to* $+\infty$,

$$\mathbf{E}\left((\hat{\theta}_j - \theta_j)|\theta_j\right) = O\left(\frac{\log n}{n}\right). \tag{9}$$

The proof of these results can be found in Section 5.

Lemma 1 is the approximation result used in Section 3 to build the estimator of the distribution of the random shifts. Lemma 2 has two consequences. First, it implies that, conditionally to θ_j, $\widehat{\theta}_j$ is an efficient estimator of θ_j at the order 1. Second, it provides an explicit form for the second order term of the quadratic risk. Thus, it extends the result of [8] to a discrete regression framework, which seems to be closer to practical applications. Note that the explicit expression of the remainder term is not needed (see Section 3) to establish the convergence rate of the plug-in estimator. Nevertheless it justifies the choice of the filter made in Section 4. Indeed, we see from (8) that an appropriate filter (h_k) is a filter such that $R^n[h, f^{[j]}]$ is as small as possible. Lemma 3 ensures that the conditional law of $\hat{\theta}_j$ is centered at θ_j, up to a $O(\log n/n)$ term.

2.3 Case of the period model

Let us now consider the period model mentioned at the end of the introduction, where symmetry of the functions is not assumed. The random variables θ_j arise this time as period of periodic functions. The observations in a fixed and equally spaced design are

$$Y_{ij} = f^{[j]}\left(\frac{i}{n\theta_j}\right) + \varepsilon_{ij} \qquad i = -n/2, \dots, n/2 \ , j = 1, \dots, J_n \ , \tag{10}$$

where the θ_j's belong to a compact interval in $]0, +\infty[$ and the 1-periodic functions $f^{[j]}$ fulfill some smoothness assumptions, for instance the ones assumed in [4] in the Gaussian white noise framework. It is established in [4] that the penalized profile likelihood method yields estimators satisfying, with appropriate rescaling, statements similar to (7) and (8), in the continuous-time model, see [4, Lemma 11 and Theorem1].

Hence one can apply the procedure presented in this paper for estimating the law of the θ_j's in model (10), provided one can transpose the proofs of (7)-(8) for the continuous-time model in terms of the discrete framework, as is done here in Section 5 for the shift model.

3 Nonparametric estimation of the distribution μ

We are interested in the estimation of the distribution of the unobserved random variables $\theta_1, \dots, \theta_{J_n}$ in the general model (1), provided identifiability holds. We shall assume that the number of curves J_n tends to $+\infty$. Our approach is based on the assumption that, along each curve, one can estimate in an appropriate way the corresponding θ_j.

More precisely, the realizations $\theta_1, \dots, \theta_{J_n}$ are unknown but we assume that they can be approximated by some preliminary estimators $\widehat{\theta}_{j,n}$ (denoted for simplicity $\widehat{\theta}_j$ in the sequel) for $j = 1, \dots, J_n$. *Approximated* means here that we assume that for each j, it is possible to build $\widehat{\theta}_j$ based on the observations $Y_{1j}, \dots, Y_{nj}$ satisfying the deviation bound (7) given in Lemma 1.

Equation (7) mainly says that each θ_j can be estimated at parametric rate with appropriate exponential control of the tail probabilities. The way of obtaining such a result in model (2) is discussed in Section 2.

3.1 A discrete estimator of μ.

A first way to define an estimator of μ is to consider a plug-in version of the usual empirical distribution, defined using the preliminary estimates $\widehat{\theta}_j$ as

$$\widehat{\mu}_{J_n} = \frac{1}{J_n}\sum_{j=1}^{J_n} \delta_{\widehat{\theta}_j}. \tag{11}$$

The empirical distribution computed with the conditional estimators of the shifts provides a consistent approximation of the distribution of the true random shifts, in a weak sense.

Theorem 1 (Weak consistency of the plugged empirical measure)**.** *Assume* (7)*, that* Θ *is compact and that there are positive finite constants* α *and* B *such that* $J_n \leqslant Bn^{\alpha}$ *and* $J_n \to +\infty$*. It holds*

$$\widehat{\mu}_{J_n} \overset{J_n\to\infty}{\rightharpoonup} \mu \quad \text{a.s.}, \tag{12}$$

which means that for all continuously differentiable compactly supported function g*,*

$$\widehat{\mu}_{J_n} g = \frac{1}{J_n}\sum_{j=1}^{J_n} g(\widehat{\theta}_j) \to \mu g = \mathbf{E}(g(\theta)) \quad \text{a.s.}$$

Proof. For g a continuously differentiable compactly supported function, we get that

$$\begin{aligned}\widehat{\mu}_{J_n} g &= \frac{1}{J_n}\sum_{j=1}^{J_n}\left(g(\widehat{\theta}_j) - g(\theta_j)\right) \quad (I)\\ &+ \frac{1}{J_n}\sum_{j=1}^{J_n} g(\theta_j) \qquad\qquad (II).\end{aligned}$$

The law of large numbers ensures that **a.s**

$$(II) \overset{J_n\to\infty}{\longrightarrow} \mathbf{E}(g(\theta)). \tag{13}$$

Now Taylor upper bound leads to $|\frac{1}{J_n}\sum_{j=1}^{J_n}\left(g(\widehat{\theta}_j) - g(\theta_j)\right)| \leqslant \frac{1}{J_n}\sum_{j=1}^{J_n}\|g'\|_\infty|\widehat{\theta}_j - \theta_j|$. If $\|g'\|_\infty = 0$, then the previous quantity is equal to zero. Now consider the case $\|g'\|_\infty \neq 0$. Hence, using prior bound and (7), we get for any $\lambda \geqslant 0$

$$\begin{aligned}\mathbf{P}\left(|\frac{1}{J_n}\sum_{j=1}^{J_n}[g(\widehat{\theta}_j) - g(\theta_j)]| \geqslant \lambda \mid \theta_1,\dots,\theta_{J_n}\right) &\leqslant \sum_{j=1}^{J_n}\mathbf{P}\left(|\widehat{\theta}_j - \theta_j| \geqslant \frac{\lambda}{\|g'\|_\infty} \mid \theta_j\right)\\ &\leqslant c_1 J_n \exp\left(-c_2\frac{\lambda^2 n\|f^{[j]}{}'\|^2}{\|g'\|_\infty^2}\right),\end{aligned}$$

which is uniform in $(\theta_j)_{j=1,\dots,J_n}$. Then, choosing $\lambda = c\sqrt{\log n/n}$ leads to the following bound.

$$\mathbf{P}\left(|\frac{1}{J_n}\sum_{j=1}^{J_n}[g(\widehat{\theta}_j) - g(\theta_j)]| \geqslant \lambda\right) \leqslant c_1 J_n n^{-\alpha} n^{\alpha - c_2c^2\|f^{[j]}{}'\|^2/\|g'\|_\infty^2}.$$

For c large enough, namely for all $\eta \geqslant 0$, $c^2 \geqslant (1+\alpha+\eta)\|g'\|_\infty^2/(c_2\|f^{[j]}{}'\|^2)$, we can write

$$\mathbf{P}\left(|\frac{1}{J_n}\sum_{j=1}^{J_n}[g(\widehat{\theta}_j) - g(\theta_j)]| \geqslant c\sqrt{\frac{\log n}{n}}\right) \leqslant c_1 n^{-(1+\eta)}.$$

Borel Cantelli's Lemma enables us to conclude that a.s.

$$\frac{1}{J_n}\sum_{j=1}^{J_n}[g(\widehat{\theta}_j)-g(\theta_j)] \overset{J_n\to+\infty}{\longrightarrow} 0. \tag{14}$$

Finally (13) and (14) prove the result. □

Hence, we have constructed a discrete estimator of the law of the random shifts. Nevertheless, in many cases this estimator is too rough when the law of the unknown effect has a density, said φ, with respect to Lebesgue measure. That is the reason why, in the following, a density estimator is built, for which we provide functional rates of convergence.

3.2 Estimation of the density of the random deformation

Consider the kernel estimator of φ based on a kernel K, to be specified in the following, and on the quantities $\hat{\theta}_j$ by

$$\hat{\varphi}(x)=\frac{1}{nh_n}\sum_{j=1}^{J_n}K\left(\frac{x-\hat{\theta}_j}{h_n}\right). \tag{15}$$

In this subsection, we shall assume that the quantities $\widehat{\theta_j}$ satisfy (7) but also the control on their expectation provided by (9). Thus the conclusion of the next theorem holds in the translation model (2), since (7) and (9) are fulfilled thanks to Lemmas 1 and 3 of the preceding section. However it might be useful in other frameworks as soon as (7)-(9) are true. For clarity in the statement of the following theorem, we shall assume that for n large enough, either $J_n \leqslant (n/\log n)^{\frac{2\beta+1}{\beta+2}}$ or the converse inequality hold, for β defined below (otherwise use a subsequence argument).

Theorem 2 (Rate of convergence of the nonparametric estimator). *Let us assume that φ is bounded and belongs to a Hölder class $H(\beta,L)$ (see [21], p.5), with $\beta>1$. Assume moreover* (7), (9) *and that the kernel K is smooth, compactly supported, of order $\lfloor\beta\rfloor$. Then the Kernel estimator $\hat{\varphi}$ defined by* (15) *achieves the following rates of convergence, as $n, J_n\to+\infty$,*

$$\sup_x \sup_{\varphi\in H(\beta,L)} \mathbf{E}\left([\hat{\varphi}(x)-\varphi(x)]^2\right)=\begin{cases} O\left(J_n^{-\frac{2\beta}{2\beta+1}}\right), & \text{if } J_n\leqslant (n/\log n)^{\frac{2\beta+1}{\beta+2}}\\ O\left((n/\log n)^{-\frac{2\beta}{\beta+2}}\right), & \text{if } J_n\geqslant (n/\log n)^{\frac{2\beta+1}{\beta+2}}\end{cases} \tag{16}$$

Thus the classical rate of convergence $J_n^{-\frac{2\beta}{2\beta+1}}$ of density estimators over Hölder classes $H(\beta,L)$, with $\beta>1$, is maintained, provided the number of curves J_n does not exceed $(n/\log n)^{\frac{2\beta+1}{\beta+2}}$, which proves optimality of the procedure in such cases. In the other cases, the number of points per curve n becomes the limiting factor, and a slower rate specified by (16) is obtained.

Hence the inverse problem is drastically reduced when the number of observations per subject increases, enabling, in a way, to invert the convolution operator. A nonparametric estimation of the density of the unobserved parameter in a regression framework can only be achieved if there are numerous observations for each curve. Indeed in our case, the asymptotics can be taken both in J_n and in n. This enables us to estimate, for each curve, the random effect and then plug the values to estimate the density. If the number of observations per curve is small, as it is usually the case in pharmacokinetics, such techniques cannot be applied and we refer to [7] for an alternative methodology.

Proof. First note that the bias-variance decomposition writes

$$\begin{aligned}\mathbf{E}\left([\hat{\varphi}(x)-\varphi(x)]^2\right) &= \left(\mathbf{E}[\hat{\varphi}(x)]-\varphi(x)\right)^2 &+& \ \mathbf{E}\left([\hat{\varphi}(x)-\mathbf{E}(\hat{\varphi}(x))]^2\right)\\ &= \qquad b(x)^2 &+& \ v(x).\end{aligned}$$

Let us denote by Δ the quantity $\hat{\theta}_1-\theta_1$. Note that, by definition of $\hat{\theta}$, Δ is a measurable function of $(\theta_1,\{\varepsilon_{i1}\}_{i=1,\dots,n})$. In the sequel, we denote $\Delta=g(\theta_1,\varepsilon)$.

Let us denote by $\mathcal{A}_1=\{|\widehat{\theta_1}-\theta_1|\leqslant D(n^{-1}\log n)^{1/2}\}$. Using (7), the probability of its complement is negligible.

A Taylor's expansion of order k of the kernel K yields the existence of a random variable Z such that

$$\begin{aligned}\frac{1}{h}\mathbf{E}K\left(\frac{\hat{\theta}_1-x}{h}\right) &= \frac{1}{h}\mathbf{E}K\left(\frac{\theta_1-x}{h}+\frac{\Delta}{h}\right)\\ &= \frac{1}{h}\mathbf{E}K\left(\frac{\theta_1-x}{h}\right) &(17)\\ &+\frac{1}{h}\mathbf{E}\left(\frac{\Delta}{h}K'\left(\frac{\theta_1-x}{h}\right)\right) &(18)\\ &+\frac{1}{h}\mathbf{E}\left(\frac{\Delta^2}{h^2}K''\left(\frac{\theta_1-x}{h}\right)\right) &(19)\\ &+\ldots+\frac{1}{h}\mathbf{E}\left(\frac{\Delta^{k-1}}{h^{k-1}}K^{(k-1)}\left(\frac{\theta_1-x}{h}\right)\right) &(20)\\ &+\frac{1}{h}\mathbf{E}\left(\frac{\Delta^k}{h^k}K^{(k)}\left(\frac{Z-x}{h}\right)\right). &(21)\end{aligned}$$

Note that, by the usual properties of a kernel of order $\lfloor\beta\rfloor$, see e.g., [21, theorem. 1.1],

$$(17)=\varphi(x)+h^\beta$$

It is assumed that the $\widehat{\theta_j}$'s satisfy (9), thus

$$\begin{aligned}(18) &= \frac{1}{h}\mathbf{E}\left(\mathbf{E}(\Delta|\theta_1)\frac{1}{h}K'\left(\frac{\theta_1-x}{h}\right)\right)\\ |(18)| &\leqslant \frac{C\log n}{nh}\int\frac{1}{h}\left|K'\left(\frac{u-x}{h}\right)\right|\varphi(u)du\\ &\leqslant \frac{C\log n}{nh}\int|K'(v)|\,\varphi(x+vh)dv\leqslant\frac{C\log n}{nh}\|\varphi\|_\infty\int|K'|.\end{aligned}$$

Splitting (19) using $\mathcal{A}_1$ and its complement,

$$|(19)|\leqslant\frac{C\log n}{nh^2}\|\varphi\|_\infty\int|K''|.$$

By the same argument,

$$|(20)| \leqslant C \sum_{p=3}^{k-1} \left(\sqrt{\frac{\log n}{nh^2}} \right)^p \leqslant \frac{C \log n}{nh^2},$$

as soon as $\log n/(nh^2) \to 0$. Finally,

$$|(21)| \leqslant \frac{C}{h} \left(\sqrt{\frac{\log n}{nh^2}} \right)^k \leqslant \frac{C}{\sqrt{nh}} \left(\frac{\log n}{n^{k-1} h^{2k+1}} \right)^{1/2}$$

Thus

$$b(x)^2 \leqslant c \left[h^{2\beta} + \frac{1}{nh} \frac{\log^2 n}{nh^3} + \frac{1}{nh} \frac{\log^k n}{n^{k-1} h^{2k+1}} \right]. \tag{22}$$

The variance term is bounded by

$$\begin{aligned}
v(x) &\leqslant \frac{1}{J_n h^2} \mathbf{E} \left(K \left(\frac{\theta_1 - x + \Delta}{h} \right)^2 \right) \\
&\leqslant \frac{1}{J_n h^2} \mathbf{E} \left[\mathbf{E} \left(K \left(\frac{\theta_1 - x + g(\theta_1, \varepsilon)}{h} \right)^2 \mid \varepsilon \right) \right] \\
&\leqslant \frac{1}{J_n h^2} \mathbf{E} \left[\int K \left(\frac{u - x + g(u, \varepsilon)}{h} \right)^2 \varphi(u) du \right] \\
&\leqslant \frac{1}{J_n h} \mathbf{E} \left[\int K \left(v - \frac{g(x + hv, \varepsilon)}{h} \right)^2 \varphi(x + hv) dv \right] \leqslant \frac{C}{J_n h} \|\varphi\|_\infty \|K\|_\infty.
\end{aligned}$$

Finally, choosing k large enough in (22), we obtain

$$\mathbf{E} \left([\hat{\varphi}(x) - \varphi(x)]^2 \right) \leqslant c \left[h^{2\beta} + \frac{1}{nh} \frac{\log^2 n}{nh^3} \right] + \frac{C}{J_n h} \|\varphi\|_\infty \|K\|_\infty \tag{23}$$

To obtain the rate of convergence of $\hat{\varphi}$, we distinguish two cases, depending on whether the second or the third term in the preceding display is dominant.

- If $J_n \leqslant (n/\log n)^{\frac{2\beta+1}{\beta+2}}$, then choosing $h_n = n^{\frac{-1}{2\beta+1}}$ implies that $\frac{1}{nh} \frac{\log^2 n}{nh^3} \leqslant \frac{C}{J_n h}$, leading to the rate
$$\mathbf{E} \left([\hat{\varphi}(x) - \varphi(x)]^2 \right) \leqslant c J_n^{-\frac{2\beta}{2\beta+1}},$$
which is the well-known optimal minimax rate when J_n observations are available.

- If $J_n \geqslant (n/\log n)^{\frac{2\beta+1}{\beta+2}}$, then choosing $h_n = (\log n/n)^{\frac{1}{\beta+2}}$, implies that $\frac{1}{nh} \frac{\log^2 n}{nh^3} \geqslant \frac{C}{J_n h}$, leading to the rate
$$\mathbf{E} \left([\hat{\varphi}(x) - \varphi(x)]^2 \right) \leqslant c n^{-\frac{2\beta}{\beta+2}}.$$
Other choices of h_n can easily be seen to lead to slower rates when optimizing (23).

□

Remark 2. Note that the difficulty of the proof relies on the fact that, a priori, $\Delta = \hat{\theta}_1 - \theta_1$ and θ_1 *are not independent*, see for instance the expression of the shift estimator given by (6). Thus one cannot easily change variables in integrals of the type $\int K\left(\frac{u-x+g(u,\varepsilon)}{h}\right)\varphi(u)du$ since g depends on u.

Remark 3. Theorem 2 requires the conditions $\beta > 1$ and (9) to be fulfilled. However, if one (or both) of these two conditions is not assumed, then it is not difficult to check from the preceding proof, using the rough bound $|(18)| \leqslant c/(\sqrt{n}h)$, that one can still recover a rate of convergence given by optimization in h of $h^{2\beta} + 1/(nh^2) + 1/(J_n h)$. This leads to a rate in $J_n^{-\frac{2\beta}{2\beta+1}}$ (respectively $n^{\frac{-\beta}{2\beta+1}}$), for J_n smaller (resp. larger) than $n^{\frac{2\beta+1}{2\beta+2}}$.

4 Simulations

In this section, we first present how the shift estimators studied in Section 2 can be numerically implemented. The estimation method proposed here is interesting on its own, since it provides a numerically tractable semiparametric estimator of the translation parameter and generalizes the penalization method proposed in [16]. Then, we construct the nonparametric estimator of the density and illustrate its behavior on both simulated data and real data.

4.1 Numerical algorithm for shift estimation and extensions

To compute explicitly $\hat{\theta}_n$ given by (6) for each curve, one has to choose an appropriate filter (h_k). This is done using a penalization technique as explained below. We refer to [16] for further references.

Consider the class of *Pinsker-type weights*, depending on the parameters K and β, defined by

$$h_k = \left[1 - (k/K)^\beta\right]_+, \quad k \geqslant 0, \tag{24}$$

and K is called the *length* of the sequence of weights.

Hence a sequence of weights is characterized by the pair (β, K). To simplify, we fix the value of β and take $\beta = 3$. Thus the family of weights depends on the single parameter K. For any filter sequence of length K, we define

$$\Lambda_K(\tau) = \sum_{k=1}^{K} h_k \left| \frac{1}{n} \sum_{i=1}^{n} \cos(2\pi k(t_i - \tau)) Y_i \right|^2, \tag{25}$$

where $Y_i,\ i = 1, \ldots, n$ is the data corresponding to one curve in model (1). To make the estimator feasible, we take the values of τ in a fixed regular grid of mesh $1/m$: $\{\tau_1, \ldots, \tau_{i+1} = \tau_1 + i/m, \ldots, \tau_{max}\}$, of range inferior to $1/2$ (let us recall that the diameter of Θ has to be bounded above by $1/2$). Let us define

$$\begin{aligned} \widehat{\tau}(K) &= \underset{\tau_1,\ldots,\tau_{max}}{\operatorname{argmax}}\ \Lambda_K(\tau), \\ M(K) &= \max_{\tau_1,\ldots,\tau_{max}} \Lambda_K(\tau). \end{aligned}$$

Penalization. We would like to find an adapted sequence of weights, or equivalently an integer K. This is done, as in [16] or [4], using a penalization method. Let

$$\widehat{K}(\alpha) = \underset{K_1,\ldots,K_{max}}{\operatorname{argmax}} \{-M(K) + \alpha K\}. \tag{26}$$

The parameter α should yield a trade-off between the fit with the data and the filter length K that can be viewed as the complexity of the chosen model. We use a data-driven method to find an appropriate α. The idea is to detect the changes in the convex hull of the function $K \to -M(K)$. Let us recall the following lemma from [16].

Lemma 4. *There exist two sequences $K_1 = 1 < K_2 < \cdots$, and $\alpha_0 = +\infty > \alpha_1 > \cdots$, with:*

$$\alpha_p = \max_{K_p < K \leqslant K_{max}} \frac{M(K_p) - M(K)}{K_p - K} = \frac{M(K_{p+1}) - M(K_p)}{K_{p+1} - K_p}, \ p \geqslant 1,$$

and such that

$$\forall\ \alpha \in (\alpha_p; \alpha_{p+1}],\ \widehat{K}(\alpha) = K_p.$$

Note that connecting the points $(K_i, M(K_i))$ gives the convex hull of the function $K \to -M(K)$ over the points $K_1, \ldots, K_{max}$. Let us define our estimator of the period as

$$\tau^* = \underset{\tau_i}{\operatorname{argmax}} \sum_{p\ :\ \widehat{\tau}(K_p) = \tau_i} \{\alpha_{p-1} - \alpha_p\}. \tag{27}$$

In words, the estimator chooses the point at which the cumulated jump in the derivative is the highest. Note that different values α^* of α led to such an estimator, which satisfies the identity $\tau^* = \hat{\tau}(\hat{K}(\alpha^*))$.

Let us note that the criterion (25) is more flexible than the one introduced in [16] since it allows weights h_k in front of the squares, weights whose asymptotic influence is made explicit by Lemma 2. The criterion used in [16] is a particular case of (25) with projection weights $h_k = \mathbf{1}_{|k| \leqslant K}$. In practice, the use of the Pinsker-type weights (24) allows a *smoothing* with respect to projection weights, making the detection of main jumps in the convex hull of $K \to -M(K)$ less sensible to local irregularities of $K \to M(K)$. A numerical comparison of the two approaches in the case of the period model in discrete time is carried out in [3].

Illustration of the algorithm. Figure 1 illustrates this algorithm with f equal to $f_1(x) = 0.015 * \cos\{100 \cos(\pi(x - \tau))\}$. The first graph represents the criterion $-M(K)$ together with, in dotted line, its convex hull. The stem diagram represents the differences $\alpha_{p-1} - \alpha_p$ corresponding to the points K_p. Finally the estimated shifts for the different values of K are represented by the last graph. The numerical parameters are the following: $n = 800$, and the true shift parameter $\tau = 0.35$. The grid for τ is the regular grid of $[0.25, 0.75]$ with 100 points. Note that, due to the high level of the noise (small amplitude of the signal with respect to the noise variance), it is difficult to detect visually the changes in the criterion behavior. Nevertheless the algorithm succeeds in finding the true shift.

The number of significative harmonics of f_1 is roughly 100, thus if we knew f_1, taking K of the order of 100 would be a reasonable choice in view of (6). In fact, for k much larger than 100, the corresponding elements in the sum of squares in (6) are mainly noise.

As we see in Figure 1, with the choice of $\widehat{\tau}$ given in (27), our algorithm chooses $\widehat{K}$ in the appropriate interval.

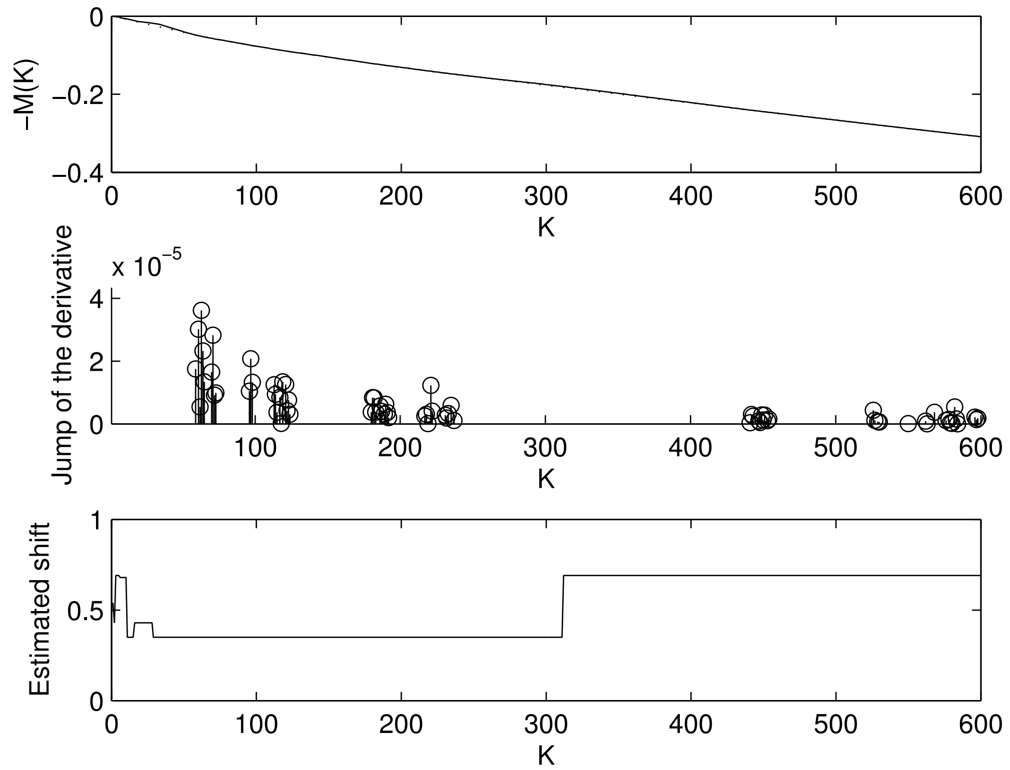

Figure 1: Finding the parameter $\widehat{\alpha}$

The period model. We note that this algorithm can also be implemented for the period model (10) and more generally if the penalized profile likelihood is known in a closed form. For the period model, the algorithm follows the description above, once one replaces the cosine in (25) by its equivalent $\cos(2\pi k(t_i/\tau))$. A numerical study is carried out in [4], leading to similar conclusions than the one presented here.

4.2 Numerical algorithm for density estimation

Once obtained the estimators of the realizations $\widehat{\theta}_j$, $j = 1, \ldots, J$, we can build the estimator of the density φ defined by (15). We illustrate the good behaviour of our algorithm with three examples. The first one shows that important features of the target density such as bimodality can be detected with our method. The second example shows that even with quite involved functions for which the semiparametric step is not easy, the methods performs well, at least if the signal to noise ratio is not too small. The third example deals with a practical application where symmetry can be seen as a sensible assumption.

Simulated data (I). The function f is the sine function on an half period, while the law of the shift θ is a bimodal Gaussian mixture. We perform 50 random translation of the original curve. In Figure 2, we present the observed curves in model (2). To study

the performance of the estimator described in this paper, we consider first preliminary estimates $\hat{\theta}_j$ obtained by the semiparametric method of Section 2, using the practical algorithm of Section 4. This set of values is used to build two nonparametric estimators of the density φ, denoted respectively *SPGaussian* and *SPepanech*, using (15) and respectively a Gaussian kernel and Epanechnikov kernel. The smoothing parameters are chosen by cross-validation.
We compare their performance with an estimate constructed the following way. Using the algorithm described in [24], applied in the shape invariant model and following the lines of Section 3.3 in [24], we compute nonparametrically the values of the warping parameter, used to align the curves to the true shape. Then, using Epanechnikov kernel, we build the corresponding density estimator, denoted by *NPplug*.

Figure 3 carries out the comparison between the preceding estimators. Visually, the estimators *SPGaussian* and *SPepanech* detect the density shape and bimodality and *SPepanech* matches slightly better the density amplitude. The nonparametric-based kernel estimator *NPplug* catches the global shape of the bimodal density but is too rough, since the method it relies on is far too general with respect to the semiparametric method designed to handle this particular situation. Hence, plugging a semiparametric preliminary estimate into a kernel-type estimator leads to a tractable estimator of the density of the shifts without knowledge of the shape of the warped function.

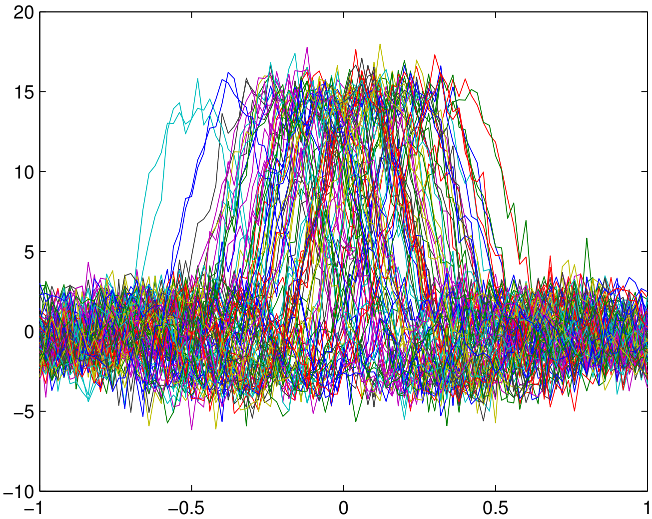

Figure 2: Simulated shifted curves.

Simulated data (II). We consider a function similar to the one introduced in the preceding subsection: $f(x) = \cos(100\cos(\pi(x-0.35)))$. Such functions appear for example in laser vibrometry and are studied in [16] and [4] for the period estimation problem. In this case, the semiparametric estimation step is crucial since the data are fuzzy. In Figure 4 we represent the original curve and both the true density and the estimated density (plotted as dotted line). We have shifted the curves using a Gaussian density, with $n = 100$ observations and $J_n = 30$. The two functions are still visually relatively close.

Real data. We present in Figure 5 an estimation conducted on real data. This data, provided by ACI-NIM MIST-R (http://www.lsp.ups-tlse.fr/Fp/Loubes/ACI.html), are daily

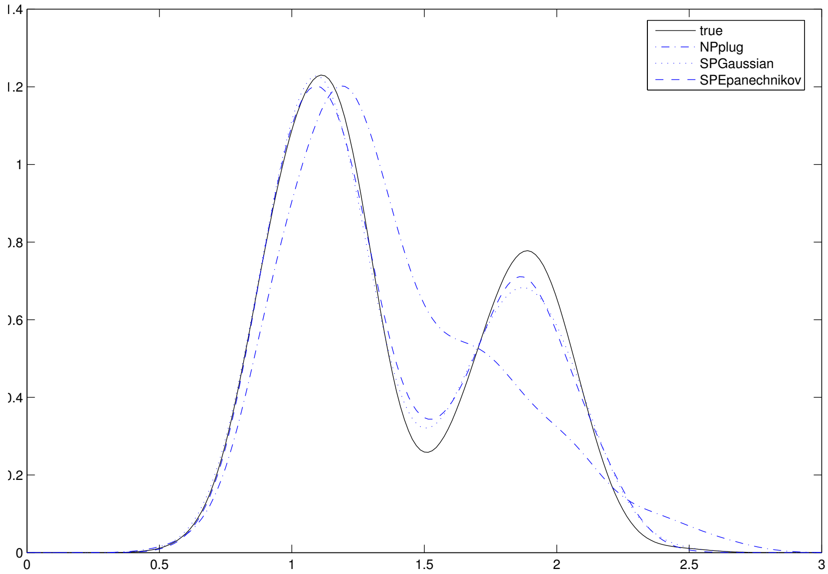

Figure 3: Estimators of the shifts density.

velocities of vehicles on a motorway on the suburbs of Paris.

After a preliminar classification which aims at building groups of homogenous curves, we obtain several functional sets, each one representing a particular daily behaviour, as pointed out in [6]. For one group we get curves starting and ending at the maximal speed, while presenting some typical patterns which stand for a standard traffic-jam feature, repeated mornings and afternoons. Due to classification, the different features have been split into different classes, as pointed out in [11]. Hence the curves present some symmetrical aspect but the starting hours of these traffic jams change slightly around a mean time, starting sooner or later each day. Hence, the shift model can be used here, as done also in [11].

In this study, we get a set of 32 curves with $n = 180$ observations which corresponds to a velocity measured every 8 minutes during a day. Understanding roadtrafficking be-

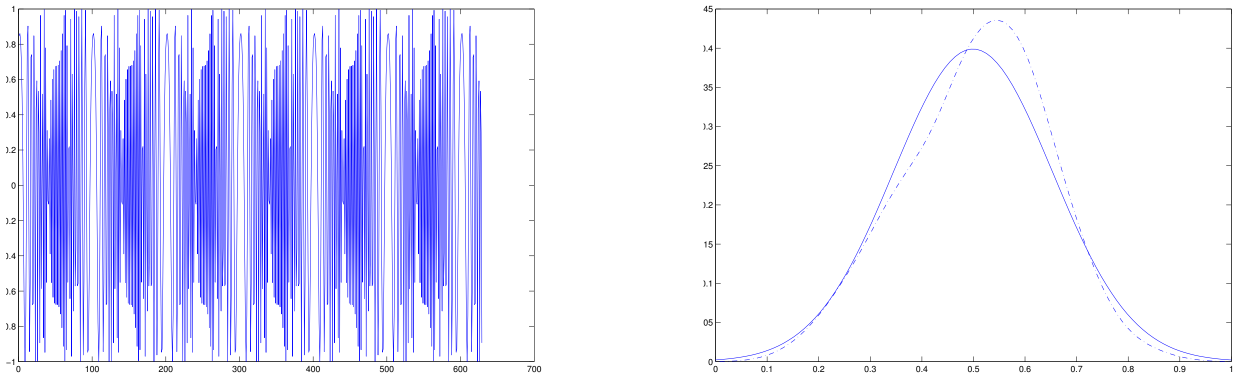

Figure 4: Simulated laser vibrometry-type functions

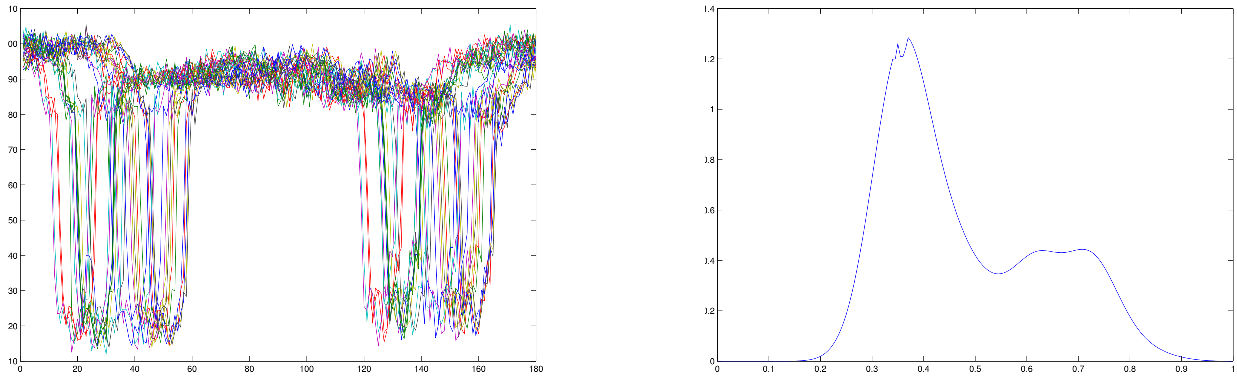

Figure 5: Real data: velocities curves

haviour, involves first finding a mean pattern but also studying the density of the random shifts, in order to understand the reasons of this changes around the mean behaviour. The bimodal feature of the estimated density can be later understood as the consequence of different weather conditions on the road network.

5 Appendix

In the proof of Lemmas 1,2 and 3, we work conditionally to the event $\theta_j = \theta$. To abbreviate the notation, we omit the index j in the proofs and drop the notation $|\theta_j)$. Thus the expectations in the following should be understood at fixed θ_j.

<u>Proof of Lemma 1</u>:

Proof. The proof follows the lines of the proof of Lemma 5 in [8]. The difference is that we have to deal with approximate Fourier coefficients instead of the f_k's. In fact, the contrast function to be maximized is

$$
\begin{aligned}
L(\tau) &= \sum_{k \geqslant 1} h_k \left[\frac{1}{n} \sum_{i=1}^{n} \sqrt{2} \cos(2\pi k(t_i - \tau)) Y_i \right]^2 \\
&= \sum_{k \geqslant 1} h_k \Big[\cos(2\pi k(\theta - \tau)) \widehat{f_k} + \sin(2\pi k(\theta - \tau)) \widehat{g_k} + \dots \\
&\quad \dots \frac{1}{\sqrt{n}} (\cos(2\pi k \tau) \xi_k + \sin(2\pi k \tau) \xi_k^*) \Big]^2
\end{aligned}
\tag{28}
$$

where

$$
\widehat{f_k} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{2} \cos(2\pi k(t_i - \theta)) f(t_i - \theta)
$$

$$
\widehat{g_k} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{2} \sin(2\pi k(t_i - \theta)) f(t_i - \theta).
$$

Note that

$$
|\widehat{f_k} - f_k| \leqslant \|f'\|_\infty / n \quad \text{and} \quad |\widehat{g_k}| \leqslant \|f'\|_\infty / n \quad \text{for all integer } k \text{ and all } \theta \text{ in } \Theta. \tag{29}
$$

$L(\tau)$ is the sum of three terms

$$L(\tau) = \eta_0(\tau) + \frac{2}{\sqrt{n}}\|f^{'}\|\eta_1(\tau) + \frac{1}{n}\eta_2(\tau),$$

where

$$\eta_0(\tau) = \sum_{k\geqslant 1} h_k[\cos(2\pi k(\theta-\tau))\widehat{f}_k + \sin(2\pi k(\theta-\tau))\widehat{g}_k]^2$$

$$\eta_1(\tau) = \|f^{'}\|^{-1}\sum_{k\geqslant 1} h_k[\cos(2\pi k(\theta-\tau))\widehat{f}_k + \sin(2\pi k(\theta-\tau))\widehat{g}_k][\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*]$$

$$\eta_2(\tau) = \sum_{k\geqslant 1} h_k[\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*)]^2.$$

Again, η_0 is a sum of three terms:

$$\begin{aligned}\eta_0(\tau) =& \sum_{k\geqslant 1} h_k\cos^2(2\pi k(\tau-\theta))\widehat{f}_k^2 + 2\sum_{k\geqslant 1} h_k\widehat{f}_k^2\widehat{g}_k^2\cos(2\pi k(\tau-\theta))\sin(2\pi k(\tau-\theta))\\ &+\sum_{k\geqslant 1} h_k\sin^2(2\pi k(\tau-\theta))\widehat{g}_k^2\\ =&\ \gamma_0(\tau)+\gamma_1(\tau)+\gamma_2(\tau).\end{aligned}$$

$$\begin{aligned}&\mathbf{P}_\theta\left(|\widehat{\theta}-\theta|\sqrt{n\|f^{'}\|^2} > x\right) \leqslant \mathbf{P}_\theta\left(\max_{\sqrt{n}\|f^{'}\||\tau-\theta|>x}(L(\tau)-L(\theta))\geqslant 0\right)\\ &\leqslant \mathbf{P}_\theta\left(\max_{\sqrt{n}\|f^{'}\||\tau-\theta|>x}\left[\eta_0(\tau)-\eta_0(\theta)+2\frac{\|f^{'}\|}{\sqrt{n}}(\eta_1(\tau)-\eta_1(\theta))+\frac{1}{n}(\eta_2(\tau)-\eta_2(\theta))\right]\geqslant 0\right)\\ &\leqslant \mathbf{P}_\theta\left(\max_{\sqrt{n}\|f^{'}\||\tau-\theta|>x}\eta_0(\tau)-\eta_0(\theta)+|\tau-\theta|\left\{2\frac{\|f^{'}\|}{\sqrt{n}}\max_{t\in\Theta}|\eta_1^{'}(t)|+\frac{1}{n}\max_{\theta\in\Theta}|\eta_2^{'}(t)|\right\}\geqslant 0\right).\end{aligned}$$

We can find $C>0$ such that $\gamma_0(\tau)-\gamma_0(\theta)\leqslant -C|\tau-\theta|^2$ for all $\tau\in\Theta$. It is the proof of Lemma 4 in [8]. Taking n large enough ($\widehat{f}_k\to f_k$ at the rate $1/n$ uniformly in k): $-2\sum_{k\geqslant 1} h_k\widehat{f}_k^2(2\pi k)^2<0$. Hence $\eta_0(\tau)-\eta_0(\theta)\leqslant -C|\tau-\theta|^2+\max_{t\in\Theta}(|\gamma_1(t)|+|\gamma_2(t)|)$.

To bound the term involving $|\eta_1'|$, we adapt the proof of Lemma 1 in [8]. Rice formula can be applied in the same way since $\mathbf{E}(\eta_1'(\tau)^2)=\|f'\|^{-2}\sum_{k\geqslant 1} h_k^2(\widehat{f}_k^2+\widehat{g}_k^2)(2\pi k)^2\geqslant c>0$ and $\mathbf{E}(\eta_1''(\tau)^2)=\|f'\|^{-2}\sum_{k\geqslant 1} h_k^2(\widehat{f}_k^2+\widehat{g}_k^2)(2\pi k)^4\leqslant c'<+\infty$ using $\widehat{f}_k^2+\widehat{g}_k^2\to f_k^2$.

The last term, involving $|\eta_2'|$, is exactly the same as in [8]: we can then use Lemma 3 there.

Now take $x=x_n=K(\log n)^{1/2}$:

$$\begin{aligned}&\mathbf{P}_\theta\left(|\widehat{\theta}-\theta|\sqrt{n\|f^{'}\|^2}>x_n\right)\\ &\quad\leqslant \mathbf{P}_\theta\left(\max_{t\in\Theta}(|\gamma_1(t)|+|\gamma_2(t)|)\geqslant C\frac{x_n^2}{n}\right)+\mathbf{P}_\theta\left(\max_{t\in\Theta}|\eta_1'(t)|\geqslant Cx_n\right)+\mathbf{P}_\theta\left(\max_{t\in\Theta}|\eta_2'(t)|\geqslant C\sqrt{n}x_n\right).\end{aligned}$$

For n large enough, the first probability above is zero since

$$\max_{t\in\Theta}(|\gamma_1(t)|+|\gamma_2(t)|)\leqslant\sum_k h_k(2|\widehat{f}_k\widehat{g}_k|+\widehat{g}_k^2)=O(n^{-1}).$$

We conclude like in [8]. □

Proof of Lemma 2:

Proof. We adapt the proof of [8] and use the notation $\|h'\| = (\sum_{k\geqslant 1}(2\pi k)^2 h_k^2)^{1/2}$. The idea is first to prove the expansion for $\widehat{\tau}$ defined by

$$L'(\theta) + (\widehat{\tau} - \theta)\mathbf{E}(L''(\theta)) = 0. \tag{30}$$

Then we prove that $\widehat{\theta}$ and $\widehat{\tau}$ are close enough so that both have the same expansion at the order 2.

For that, we compute the derivatives of the criterion L using (29).
Let $\xi_k(\theta) = n^{-1}\sum_{i=1}^n \sqrt{2}\cos(2\pi k(t_i-\theta))\varepsilon_i$ and $\xi_k^*(\theta) = n^{-1}\sum_{i=1}^n \sqrt{2}\sin(2\pi k(t_i-\theta))\varepsilon_i$.

$$L'(\theta) = 2\textstyle\sum_k h_k(2\pi k)[-\widehat{g_k} + n^{-1/2}\xi_k^*(\theta)][\widehat{f_k} + n^{-1/2}\xi_k(\theta)] \tag{31}$$

$$L''(\theta) = -2\textstyle\sum_k h_k(2\pi k)^2[\widehat{f_k}^2 + 2n^{-1/2}\widehat{f_k}\xi_k(\theta) + \widehat{g_k}^2 + n^{-1}(\xi_k(\theta)^2 - \xi_k^*(\theta)^2)] \tag{32}$$

$$L^{(3)}(\tau) = -8\textstyle\sum_k h_k(2\pi k)^3[n^{-1}\sum_i\cos(2\pi k(t_i-\tau))Y_i][n^{-1}\sum_i \sin(2\pi k(t_i-\tau))Y_i]. \tag{33}$$

Using (29) together with (31),(32), we obtain

$$\mathbf{E}(L'(\theta)^2) = 4n^{-1}\sum_k h_k^2(2\pi k)^2(\widehat{f_k}^2 + n^{-1}) + O(n^{-2}),$$

$$\mathbf{E}(L''(\theta))^2 = 4\left[\|f'\|^2 + \sum_k (h_k-1)(2\pi k)^2 f_k^2 + \sum_k h_k(2\pi k)^2(\widehat{f_k}^2 - f_k^2 + g_k^2)\right]^2.$$

Now remark that using Cauchy-Schwarz, (29) and (A3),
$\sum_k h_k(2\pi k)^2(\widehat{f_k}^2 - f_k^2) \leqslant (\sum h_k^2)^{1/2}(\sum(2\pi k)^4(\widehat{f_k}^2 - f_k^2)^2)^{1/2} \leqslant \|h'\|O(n^{-1})$. Note that thanks to (C2), $\|h'\| = o(\|h'\|^2)$ and thus $\sum_k h_k(2\pi k)^2(\widehat{f_k}^2 - f_k^2) = o(n^{-1}\|h'\|^2)$. Hence with $I_n(f) = n\|f'\|^2$,

$$\mathbf{E}((\widehat{\tau}-\theta)^2 I_n(f)) = \left[1 + \|f'\|^{-2}\sum(h_k^2-1)(2\pi k)^2 f_k^2 + \|f'\|^{-2}\sum(h_k^2-1)(2\pi k)^2(\widehat{f_k}^2 - f_k^2)\right]\times$$
$$\left[1 - \|f'\|^{-2}\sum(1-h_k)(2\pi k)^2 f_k^2 + o(n^{-1}\|h'\|^2)\right]^{-2}.$$

Using Taylor formula at the order 2, we obtain that the second term of the previous product is equal to $1 + 2\|f'\|^{-2}\sum_k(1-h_k)(2\pi k)^2 f_k^2 + o(n^{-1}\|h'\|^2) + 3\|f'\|^{-4}(\sum_k(1-h_k)(2\pi k)^2 f_k^2)^2 + o(n^{-1}\|h'\|^2)$. Hence calculating the product

$$\mathbf{E}((\widehat{\tau}-\theta)^2 I_n(f)) = 1 + (1+o(1))\|f'\|^{-2}R_n(h,f).$$

Let us now control the difference $(\widehat{\theta} - \widehat{\tau})$. It is sufficient to do this on the set $\mathcal{A}_1 = \{|\widehat{\theta} - \theta| \leqslant D(n^{-1}\log n)^{1/2}\}$ since the probability of its complement is negligible thanks to Lemma 1.

By definition of $\hat{\theta}$, $L'(\hat{\theta}) = 0$. By Taylor's expansion, for some random variable ω between θ and $\hat{\theta}$,

$$0 = L'(\hat{\theta}) = L'(\theta) + (\hat{\theta}-\theta)L''(\theta) + \frac{(\hat{\theta}-\theta)^2}{2}L^{(3)}(\omega),$$

which can also be written

$$\begin{aligned} 0 &= L'(\theta) + (\hat{\theta} - \theta)\mathbf{E}(L''(\theta)) \\ &\quad + (\hat{\theta} - \theta)[L''(\theta) - \mathbf{E}(L''(\theta))] + \frac{(\hat{\theta} - \theta)^2}{2} L^{(3)}(\omega). \end{aligned} \tag{34}$$

Substracting (30) and (34), one obtains

$$\mathbf{E}((\widehat{\theta} - \widehat{\tau})^2 \mathbf{1}_{\mathcal{A}_1}) \leqslant \mathbf{E}(L''(\theta))^{-2}[2\mathbf{E}((\widehat{\theta} - \theta)^2 (L''(\theta) - \mathbf{E}(L''(\theta)))^2 \mathbf{1}_{\mathcal{A}_1}) + \mathbf{E}((\widehat{\theta} - \theta)^4 \sup_{\omega \in \Theta} |L^{(3)}(\omega)|^2)].$$

Easy calculations lead to $n\mathbf{E}\left[L''(\theta) - \mathbf{E}(L''(\theta))\right]^2 = 16 \sum_k h_k^2 (2\pi k)^4 (\widehat{f_k}^2 + n^{-1})$ and thus this quantity is bounded above by a constant thanks to (A3) and (C3). Similarly, following the proof of Lemma 6 in [8] and using (33), one obtains $|L^{(3)}(\omega)| \leqslant C \sum_k h_k (2\pi k)^3 (|\widehat{f_k}|^2 + |\widehat{g_k}|^2 + n^{-1}(\xi_k^2 + \xi_k^{*2}))$ for all ω in Θ. Hence with (A3) and (C3), one obtains $\mathbf{E}(\sup_{\omega \in \Theta} |L^{(3)}(\omega)|^2) \leqslant C$. Thus $\mathbf{E}((\widehat{\theta} - \widehat{\tau})^2 I_n(f) \mathbf{1}_{\mathcal{A}_1}) \leqslant C n^{-1} \log^2 n$ is a $o(R_n(h, f))$.

By similar arguments, one also sees that $\mathbf{E}((\widehat{\theta} - \widehat{\tau})(\widehat{\tau} - \theta)\mathbf{1}_{\mathcal{A}_1})$ is a $o(R_n(h, f))$. □

Proof of Lemma 3:

Proof. By the triangle inequality, the Cauchy-Schwarz inequality and (34),

$$\begin{aligned} \left|\mathbf{E}\left((\hat{\theta} - \theta)\right)\mathbf{E}(L''(\theta))\right| &\leqslant |\mathbf{E}(L'(\theta))| \\ &\quad + (\mathbf{E}(\hat{\theta} - \theta)^2)^{1/2} \left\{\mathbf{E}[L''(\theta) - \mathbf{E}(L''(\theta))]^2\right\}^{1/2} \\ &\quad + \frac{1}{2}\mathbf{E}\left\{(\hat{\theta} - \theta)^2 \sup_{\omega \in \Theta} |L^{(3)}(\omega)|\right\}. \end{aligned}$$

Thanks to (31) and (29),

$$\begin{aligned} \mathbf{E}(L'(\theta)) &= \mathbf{E}\left[2\sum_k h_k(2\pi k)[-\widehat{g_k} + n^{-1/2}\xi_k^*(\theta)][\widehat{f_k} + n^{-1/2}\xi_k(\theta)]\right] \\ &= -2\sum_k h_k(2\pi k)\widehat{f_k}\widehat{g_k} = -2\sum_k h_k(2\pi k)(f_k + O(n^{-1}))O(n^{-1}) \\ &= O(n^{-1}). \end{aligned}$$

Besides, from the proof of Lemma 2 and the result of Lemma 2 respectively, we know that

$$\mathbf{E}[L''(\theta) - \mathbf{E}(L''(\theta))]^2 = O(n^{-1})$$

and

$$\mathbf{E}(\hat{\theta} - \theta)^2 = O(n^{-1}).$$

Working on the event $\mathcal{A}_1 = \{|\widehat{\theta} - \theta| \leqslant D(n^{-1}\log n)^{1/2}\}$ and using (see the proof of Lemma 2) the fact that $\mathbf{E}(\sup_{\omega \in \Theta} |L^{(3)}(\omega)|^2) \leqslant C$, it is also easy to see that

$$\mathbf{E}\left\{(\hat{\theta} - \theta)^2 \sup_{\omega \in \Theta} |L^{(3)}(\omega)|\right\} = O\left(\frac{\log n}{n}\right).$$

The fact that $C\|f'\|^2 \leqslant \mathbf{E}(L''(\theta)) \leqslant C'\|f'\|^2$ yields the result. □

Acknowledgments. The authors are grateful to Professor Alexandre Tsybakov for an insightful remark concerning an earlier version of this work.

References

[1] D. Bates and M. Lindstrom, Nonlinear mixed effects models for repeated measures data, *Biometrics*,(46), 3, 673–687, 1990.

[2] L. Brumback and M. Lindstrom Self modeling with flexible, random time transformations, *Biometrics*, (60), 2, 461–470, 2004.

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

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

[5] L. Cavalier, G.K. Golubev, D. Picard, and A.B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Stat.*, 30(3):843–874, 2002.

[6] Loubes, Jean-Michel, Maza, Élie, Lavielle, Marc, Rodríguez, Luis Road trafficking description and short term travel time forecasting, with a classification method, *Canad. J. Statist.*, 34(3), 475–491, 2006.

[7] D. Chafai and J-M. Loubes. Maximum likelihood for a certain class of inverse problems: an application to pharmakocinetics. *SPL*, 76:1225–1237, 2006.

[8] A. S. Dalalyan, G. K. Golubev, and A. B. Tsybakov. Penalized Maximum Likelihood and Semiparametric Second Order Efficiency, *Ann. Stat.*, 34(1):169–201, 2006.

[9] S. Darolles, J-P. Florens, and E. Renault. Nonparametric instrumental regression. *to appear in Econometrica*, 2005.

[10] M. Davidian and D. Giltinan. Nonlinear Models for Repeated Measurement Data: An Overview and Update. *Journal of Agricultural, Biological, and Environmental Statstics*, 8:387–419, 2003.

[11] F. Gamboa and J-M. Loubes and E. Maza. Shifts estimation with M-estimators. to appear in *Electronic Journal of Statistics*, 2007.

[12] T. Gasser and A. Kneip. Searching for structure in curve samples. *J. Amer. Statist. Assoc.*, 90:1179–1188, 1995.

[13] D. Gervini and T. Gasser. Self-modelling warping functions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(4):959–971, 2004.

[14] A. Kneip and T. Gasser. Statistical tools to analyze data representing a sample of curves. *Ann. Statist.*, 20(3):1266–1305, 1992.

[15] A. Kneip an X. Li and B. MacGibbon and J. Ramsay. Curve Registration by Local Regression. *Canadian Journal of Statistics*, 28:19–30, 2000.

[16] M. Lavielle and C. Lévy-Leduc. Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals, *IEEE Transactions on Signal Processing*, 53(7): 2306–2315, 2005.

[17] B. Lindsay. The geometry of mixture likelihoods: a general theory. *Ann. Statist.*, 11(1):86–94, 1983.

[18] F. Mentré and A. Mallet. Handling covariates in population pharmacokinetics. *Int. J. Biomed. Comp.*, 36:25–33, 1994.

[19] J. Ramsay and X. Li. Curve registration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(2):351–363, 1998.

[20] B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B.*, 69(2):243–259, 2001.

[21] A. B. Tsybakov. *Introduction à l'estimation non-paramétrique. (Introduction to non-parametric estimation).* Mathématiques & Applications (Paris). 41. Paris: Springer, 2004.

[22] A.W. van der Vaart. *Asymptotic statistics.* Cambridge University Press, 1998.

[23] M. Vimond. *Asymptotic efficiency of shifts estimators with M-estimators* to appear in Annals. of Stat, 2008.

[24] K. Wang and T. Gasser. Synchronizing sample curves nonparametrically. *Ann. Statist.*, 27(2):439–460, 1999.