

Estimation of the distribution of random shifts deformation

I. CASTILLO & J-M. LOUBES

Abstract

Consider discrete values of functions shifted by unobserved translation effects, which are independent realizations of a random variable with unknown distribution μ , modeling 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. Building on results of Dalalyan et al. (2006), semiparametric estimators are obtained in our discrete framework and their performance studied. From these estimates we construct a nonparametric estimator of the target density. Both rates 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 common density φ of independent random variables θ_j , $j = 1, \dots, J_n$, with distribution μ , observed in a panel data analysis framework in a translation model. More precisely, consider J_n unknown curves $t \rightarrow f^{[j]}(t)$ sampled at multiple points $t_{ij} = t_i = i/n$, $i = 1, \dots, n$, with random i.i.d. translation effects θ_j , $j = 1, \dots, J_n$, in the following regression framework

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

where ε_{ij} are independent standard normal $\mathcal{N}(0, 1)$ random noise and are independent of the θ_j 's, while σ is a positive real number which is assumed to be known. The number of points per curve is denoted by n while J_n stands for the number of curves.

Equation (1) describes the situation often encountered in biology, data mining or econometrics (see e.g [20] or [6]) where the outcome of an experiment depends on a random variable θ which models the case where the data variations take into account the variability of each individual: each subject j can react in a different way within a mean behaviour, with slight variations given by the unknown curves $f^{[j]}$. Estimating φ , the density of the unobserved θ_j 's, enables to understand this mean behaviour.

Nonparametric estimation of φ belongs to the class of inverse problems for which the subject of the inversion is a probability measure, since the realizations θ_j are warped

by unknown functions $f^{[j]}$'s. Here, these functions are unknown, hence the underlying inverse problem becomes more than harmful as sharp approximations of the θ_j 's are needed to prevent 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.

Note that estimation of such warping parameters have been investigated by several authors using nonparametric methods for very general models, see for instance [12, 14], or [24]. However little attention is paid to the law of these random parameters. Moreover, as said previously, sharp estimates of the parameters are required to achieve density estimation, which requires semiparametric methods.

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 [7] in the Gaussian white noise framework and extend it to the discrete 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 preliminary 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 significantly 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 [15] 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 for instance [15]. Given a sequence of J_n experiments like the one considered in [15], one might be interested in estimating the law of the corresponding periods of the signals. In this case one can also consider applying our method, under some conditions made explicit in the sequel.

The paper falls into the following parts. In Section 2, semiparametric estimators $\widehat{\theta}_j$ of the realizations of the shift parameters are proposed, and sharp bounds between $\widehat{\theta}_j$ 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. 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 (1). To build this estimates, we follow the method introduced by Dalalyan, Golubev and Tsybakov in [7] for a continuous-time version of the translation model. We obtain analogues of two of their results in our discrete-time model: a deviation estimate stated in Lemma 2.3 and a second order expansion for the estimators stated in Lemma 2.4. We also establish a new result in Lemma 2.5,

which enables to control the bias of the estimates. The particular form of the estimators and the second order expansion will not be used to build the density estimator in Section 3.

Hence, conditionally to the event $\theta_j = \theta$, we construct an estimator $\widehat{\theta}_j$ and establish asymptotic results for the conditional distribution $(\widehat{\theta}_j \mid \theta_j = \theta)$, gathered in Lemmas 2.3, 2.4 and 2.5. 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 $\|\cdot\|$ the L^2 -norm on $[0, 1]$ and by $\|\cdot\|_\infty$ the L^∞ -norm on \mathbb{R} .

2.1 Shift estimation in the discrete time translation model

The model reduces to, assuming for simplicity that $\sigma = 1$,

$$Y_i = f(t_i - \theta) + \varepsilon_i \quad i = 1, \dots, n, \quad (2)$$

where f is a *symmetric* function satisfying some additional assumptions detailed below and $t_i = i/n$. 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 compact and contained in an interval of diameter upper-bounded by $1/2$

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

The function f is assumed to be symmetric (that is, $f(x) = f(-x)$ for all real x) and periodic with period 1 with Fourier coefficients denoted by f_k , $k \geq 1$,

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

Let $\mathcal{C}^2(\mathbb{R})$ denote the set of all twice continuously differentiable functions on \mathbb{R} . We assume that there exist $\rho > 0$ and $C_0 < +\infty$ such that f belongs to the set F defined by

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

Conditions (A1)-(A3) can be seen as working assumptions and are essentially the same as in [7]. 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 belongs to $\mathcal{C}^2(\mathbb{R})$ is handy in particular for proving the second order properties of the estimator. Note also that for simplicity in the definitions of the classes, as in [7] we have assumed that the Fourier coefficient for $k = 0$, that is $\int_0^1 f(u) du$, is zero.

Identifiability in model (2) follows from : symmetry, 1-periodicity of the functions (note that assuming that $f_1^2 \geq \rho$ 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 define an estimator $\hat{\theta}$ of the shift θ in model (2) by

$$\hat{\theta} = \operatorname{argmax}_{\tau \in \Theta} \sum_{k \geq 1} h_k \left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi k(t_i - \tau)) Y_i \right)^2, \quad (3)$$

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 $\hat{\theta}$ is similar to the estimator $\hat{\theta}_{PML}$ proposed in [7]: here the integral in their definition is replaced by the equivalent discrete sum in the discrete-time model. As we sketch below, the estimator (3) arises in a natural way by using a penalized profile likelihood method as in [7], though here in an approximate way only,

First one turns the study of the regression model into the study of a sequence of independent submodels. Let us introduce, for any $k \geq 1$,

$$\begin{aligned} x_k &= \frac{1}{n} \sum_{i=1}^n \sqrt{2} \cos(2\pi k t_i) Y_i, & \xi_k &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sqrt{2} \cos(2\pi k t_i) \varepsilon_i. \\ x_k^* &= \frac{1}{n} \sum_{i=1}^n \sqrt{2} \sin(2\pi k t_i) Y_i, & \xi_k^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sqrt{2} \sin(2\pi k t_i) \varepsilon_i. \end{aligned}$$

Note that x_k and x_k^* are observed. Using the fact that Y_i follows (2),

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

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

where $d_{k,n}, d_{k,n}^*$ 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. We would like to underline two important facts about the previous quantities. First, since the ε_i 's are Gaussian $\mathcal{N}(0, 1)$, the variables $(\xi_k, \xi_k^*)_{k \geq 1}$ are also Gaussian and, since we assume that $t_i = i/n$, using the orthogonality of the trigonometric basis over this system of points, they are in fact independent standard Normal. Second, it is important to note that both $d_{k,n}$ and $d_{k,n}^*$ are non-random and bounded uniformly in θ . We will use more precise bounds in function of k and n in the proofs, see Lemma 5.1 in Section 5.

The penalized profile likelihood method is as follows. For each integer k and $\tau \in \Theta$, let us define the quantity $p_\tau(x_k, x_k^*, f_k)$ as

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

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. The profile likelihood technique

(see [22, Chap. 25]), consists in "profiling out" the nuisance parameter f_k by setting

$$f_k^*(\tau) = \operatorname{argmax}_{f_k} p_\tau(x_k, x_k^*, f_k)$$

$$\widehat{\theta}_{PML} = \operatorname{argmax}_{\tau \in \Theta} \prod_{k \geq 1} p_\tau(x_k, x_k^*, f_k^*(\tau)).$$

Here a difficulty is that $d_{k,n}$ and $d_{k,n}^*$ depend on f_k . However, if we neglect those terms, we can follow the calculations made in [7] and we obtain that $\widehat{\theta}_{PML}$ is the maximizer of

$$\sum_{k \geq 1} h_k \left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi k \tau) x_k + \sin(2\pi k \tau) x_k^* \right)^2,$$

which is exactly the same as (3) if we set $h_k = \sigma_k^2 / (\sigma_k^2 + n^{-1})$. Thus the criterion (3) can be obtained by an *approximate* profile likelihood method. Yet, it is not trivial to see at this point if having neglected the terms $d_{k,n}, d_{k,n}^*$ in the construction of the estimator can have a negative influence over the behavior of the criterion (3). In fact, we will see in the sequel that this is not the case, and that (3) can lead to a very good estimator, even at second order, provided a sensible choice of (h_k) is made.

2.2 Asymptotic behavior of the shifts estimators

First let us precise some technical assumptions we make on the sequence of weights (h_k) in the definition (3) of the estimator. These are the same as Assumptions B and C in [7], except that here we also restrict ourselves to a finite number of nonzero weights.

Assumptions on the sequence of weights (h_k)

Let the sequence (h_k) be such that $h_1 = 1$, $0 \leq h_k \leq 1$ for all $k \geq 1$ and assume that 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) $\left[\sum_{k \geq 1} (2\pi k)^2 h_k^2 \right]^{1/2} \geq \rho_1 (\log^2 n) \max_{k \geq 1} (2\pi k) h_k.$

(C3) $\sum_{k \geq 1} h_k (2\pi k)^4 \leq D_1 n.$

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

The first condition is quite natural to make the estimator feasible in practice, 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 2.4.

Remark 2.1. As noted in [7], conditions (C1), (C2), (C3) and (T) are fulfilled for a quite wide range of weights. For instance, the sequences $(h_k = \mathbf{1}_{1 \leq k \leq N(T)})$, also called projection weights, satisfy the preceding conditions since (C2) and (C3) are satisfied respectively for $N(T) \geq C \log^4 n$ and $N(T) \leq C n^{1/5}$, while condition (T) is always satisfied for projection weights since $\sum_{k \geq N(T)} (2\pi k)^2 f_k^2 \rightarrow 0$, as $n \rightarrow +\infty$, due to (A3).

Remark 2.2. It is also possible to consider random, data-driven, weights. This approach is considered in [8].

Asymptotic properties

For easiness of reference in Section 3, it is convenient here to make the dependence in j explicit again. In the following Lemmas, $f^{[j]}'$ and $f_k^{[j]}$ respectively denote the derivative and the Fourier coefficients of $f^{[j]}$. Lemmas 2.3 and 2.4 are respectively analogues of Lemma 5 and Theorem 1 in [7] here in a discrete regression framework, which seems to be closer to practical applications. Though it seems natural that the cited results extend to our context, it is not obvious that the extra terms induced by the discretization of the model, for instance the $d_{n,k}$'s introduced above, do not interfere with the rates, in particular at the second order. But we prove that they eventually do not, see the proof of the Lemmas in Section 5.

Lemma 2.3 (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*

$$\mathbf{P} \left(\sqrt{n} |\hat{\theta}_j - \theta_j| > x_n |\theta_j| \right) \leq c_1 \exp(-c_2 x_n^2). \quad (6)$$

Lemma 2.4 (Second order Expansion). *Assume that (A), (C), (T) are fulfilled. Let us denote $R^n[h, f^{[j]}] = \sum_{k=1}^{\infty} (2\pi k)^2 [(1 - h_k)^2 f_k^{[j]2} + h_k^2/n]$. Uniformly in $j \in \{1, \dots, J_n\}$, $\theta_j \in \Theta$ and $f^{[j]} \in F$, as n tends to $+\infty$,*

$$\mathbf{E} \left((\hat{\theta}_j - \theta_j)^2 | \theta_j \right) = \frac{1}{n \|f^{[j]}'\|^2} \left(1 + (1 + o(1)) \frac{R^n[h, f^{[j]}]}{\|f^{[j]}'\|^2} \right). \quad (7)$$

Lemma 2.4 implies that, conditionally to θ_j , the estimator $\hat{\theta}_j$ is efficient for estimating θ_j at the first order. It also provides an explicit form for the second order term of the quadratic risk. The explicit expression of this term is not needed to establish the convergence rate of the plug-in estimator in Section 3. Nevertheless it justifies the choice of the filter made in Section 4. Indeed, we see from (7) that an appropriate filter (h_k) is a filter such that $R^n[h, f^{[j]}]$ is as small as possible.

The following Lemma 2.5 is new with respect to [7]. It ensures that the conditional law of $\hat{\theta}_j$ is centered at θ_j , up to a $O(\log n/n)$ term.

Lemma 2.5 (Asymptotical Bias). *Assume (A), (C), (T) and that there exists a positive constant D such that for any $f \in F$, it holds $\sum_{k \geq 1} k^2 |f_k| \leq D$. Then, uniformly in $j \in \{1, \dots, J_n\}$, as n tends to $+\infty$,*

$$\mathbf{E} \left((\hat{\theta}_j - \theta_j) | \theta_j \right) = O \left(\frac{\log n}{n} \right). \quad (8)$$

This Lemma requires only slightly more regularity on f than a second derivative bounded in L^2 , which is what condition (A3) imposes, that is $\sum_{k \geq 1} k^4 f_k^2 \leq C_0$. It enables us to have slightly broader framework for our results in Section 3. However, one can still obtain interesting results without using this Lemma, see Remark 3.5 after Theorem 3.3.

2.3 Case of the period model

Let us now consider the period model mentioned in 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

$$Y_{ij} = f^{[j]} \left(\frac{i}{n\theta_j} \right) + \varepsilon_{ij} \quad i = -n/2, \dots, n/2, j = 1, \dots, J_n, \quad (9)$$

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 (6) and (7), in the continuous-time model, see [4, Lemma 11 and Theorem 1].

Hence one can apply the method of this paper for estimating the law of the θ_j 's in model (9), provided one can transpose the proofs of (6)-(7) 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 sample $\theta_1, \dots, \theta_{J_n}$ in model (1). 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 $\hat{\theta}_{j,n}$, for $j = 1, \dots, J_n$, denoted for simplicity $\hat{\theta}_j$ in the sequel.

Definition 3.1. We say that the random variables $\hat{\theta}_1, \dots, \hat{\theta}_{J_n}$ approximate the sample $\theta_1, \dots, \theta_{J_n}$ if for each j , the variable $\hat{\theta}_j$ is built using the observations Y_{1j}, \dots, Y_{n_j} (i.e is measurable with respect to these observations) and satisfies the deviation bound (6) given in Lemma 2.3.

Note in particular that with this definition, the random variables $\hat{\theta}_j$ are independent. The fact that (6) holds roughly means that the θ_j 's are approximated by the $\hat{\theta}_j$'s at almost parametric rate with an exponential control of the deviation probability.

In Section 2, we have studied in details a possible way of obtaining $\hat{\theta}_j$'s satisfying this approximation property in model (1). However, we would like to point out that the results of Section 3 hold as long as the $\hat{\theta}_j$'s are approximations of the θ_j 's in model (1) in the sense of Definition 3.1 (for Theorem 3.3 below, we shall also assume that (8) holds), which possibly allows using estimators produced by other methods. Extensions to frameworks beyond model (1) could also be considered.

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 $\hat{\theta}_j$ as

$$\hat{\mu}_{J_n} = \frac{1}{J_n} \sum_{j=1}^{J_n} \delta_{\hat{\theta}_j}. \quad (10)$$

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 3.2 (Weak consistency of the plugged empirical measure). *Assume (6), that Θ is compact and that there are positive finite constants α and B such that $J_n \leq Bn^\alpha$ and $J_n \rightarrow +\infty$. Then it holds*

$$\widehat{\mu}_{J_n} \xrightarrow{J_n \rightarrow \infty} \mu \quad \text{almost surely,} \quad (11)$$

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

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

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

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

The law of large numbers ensures that, almost surely,

$$(II) \xrightarrow{J_n \rightarrow \infty} \mathbf{E}(g(\theta)). \quad (12)$$

Now Taylor upper bound leads to $|\frac{1}{J_n} \sum_{j=1}^{J_n} (g(\widehat{\theta}_j) - g(\theta_j))| \leq \frac{1}{J_n} \sum_{j=1}^{J_n} \|g'\|_\infty |\widehat{\theta}_j - \theta_j|$. 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 (6), we get for any $\lambda \geq 0$

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

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

$$\mathbf{P} \left(\left| \frac{1}{J_n} \sum_{j=1}^{J_n} [g(\widehat{\theta}_j) - g(\theta_j)] \right| \geq \lambda \right) \leq c_1 J_n n^{-\alpha} n^{\alpha - c_2 c^2 \|f^{[j]'}\|^2 / \|g'\|_\infty^2}.$$

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

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

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

$$\frac{1}{J_n} \sum_{j=1}^{J_n} [g(\widehat{\theta}_j) - g(\theta_j)] \xrightarrow{J_n \rightarrow +\infty} 0. \quad (13)$$

Finally (12) and (13) prove the result. \square

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's 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 following kernel estimator of φ , the density of the unobserved θ_j 's in model (1), based on a kernel K , to be specified in the following, and on the quantities $\hat{\theta}_j$. For all x in Θ , let us define

$$\hat{\varphi}(x) = \frac{1}{J_n h_n} \sum_{j=1}^{J_n} K\left(\frac{x - \hat{\theta}_j}{h_n}\right). \quad (14)$$

In this subsection, we shall assume that the quantities $\hat{\theta}_j$ satisfy the approximation property stated in Definition 3.1 and also the control on their expectation provided by (8). We have checked in Section 2 that both properties are fulfilled under some regularity conditions for the estimators $\hat{\theta}_j$ built in Section 2.1.

Let us denote by $H_M(\beta, L)$ the set of all densities φ with support included in the interval $[-\tau_0, \tau_0] = \Theta$, which belong to the Hölder class $H(\beta, L)$ (see [21], p.5) and are uniformly bounded by a positive constant M .

For clarity in the following statement, we shall assume that for n large enough, either $J_n \leq (n/\log n)^{\frac{2\beta+1}{\beta+2}}$ or the converse inequality hold, for β defined below (otherwise use a subsequence argument).

Theorem 3.3 (Rate of convergence of the nonparametric estimator). *Let us assume that φ belongs to the class $H_M(\beta, L)$ for some positive L and M and with $\beta > 1$. Assume moreover (6), (8) and that the kernel K is smooth, compactly supported, of order $\lfloor \beta \rfloor$. Then the kernel estimator $\hat{\varphi}$ defined by (14) achieves the following rates of convergence, as n and J_n tend to $+\infty$,*

$$\sup_{x \in \Theta} \sup_{\varphi \in H_M(\beta, L)} \mathbf{E} ([\hat{\varphi}(x) - \varphi(x)]^2) = \begin{cases} O\left(J_n^{-\frac{2\beta}{2\beta+1}}\right), & \text{if } J_n \leq (n/\log n)^{\frac{2\beta+1}{\beta+2}} \\ O\left((n/\log n)^{-\frac{2\beta}{\beta+2}}\right), & \text{if } J_n \geq (n/\log n)^{\frac{2\beta+1}{\beta+2}} \end{cases} \quad (15)$$

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 $(n/\log n)^{\frac{2\beta+1}{\beta+2}}$. In fact, it can be checked, using standard lower bound techniques, that in model (1), the minimax rate of convergence of the pointwise mean-squared risk for the estimation of φ over the considered Hölder-class is not faster than constant times $J_n^{-\frac{2\beta}{2\beta+1}}$, which yields the rate-optimality of the procedure in this model in the first case of the Theorem. In the other case, the number of points per curve n becomes the limiting factor, and a slower rate specified by (15) is obtained. Whether this second rate is optimal is an open question.

Our results can be interpreted as follows. 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. In our case, the fact that asymptotics can be taken both in J_n and in n enables us to estimate first, 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 [6] for an alternative methodology.

Proof. For simplicity in the notation, we assume throughout the proof that the $\hat{\theta}_j$'s are identically distributed - let us recall that they are independent -, which enables us to just deal with $j = 1$. If this is not the case, then one can still use the independence and then bound the different quantities arising j by j . Also we denote h_n simply by h .

First note that the bias-variance decomposition is, for any x in Θ ,

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

Let us denote by Δ the quantity $\hat{\theta}_1 - \theta_1$. Note that, by definition of $\hat{\theta}_1$, 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 $\mathcal{A}_1 = \{|\hat{\theta}_1 - \theta_1| \leq D(n^{-1} \log n)^{1/2}\}$. Using (6), the probability of its complement is negligible.

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

$$\begin{aligned} \frac{1}{h} \mathbf{E} K \left(\frac{\hat{\theta}_1 - x}{h} \right) &= \frac{1}{h} \mathbf{E} K \left(\frac{\theta_1 - x}{h} + \frac{\Delta}{h} \right) \\ &= \frac{1}{h} \mathbf{E} K \left(\frac{\theta_1 - x}{h} \right) \end{aligned} \tag{16}$$

$$+ \frac{1}{h} \mathbf{E} \left(\frac{\Delta}{h} K' \left(\frac{\theta_1 - x}{h} \right) \right) \tag{17}$$

$$+ \frac{1}{h} \mathbf{E} \left(\frac{\Delta^2}{h^2} K'' \left(\frac{\theta_1 - x}{h} \right) \right) \tag{18}$$

$$+ \dots + \frac{1}{h} \mathbf{E} \left(\frac{\Delta^{k-1}}{h^{k-1}} K^{(k-1)} \left(\frac{\theta_1 - x}{h} \right) \right) \tag{19}$$

$$+ \frac{1}{h} \mathbf{E} \left(\frac{\Delta^k}{h^k} K^{(k)} \left(\frac{Z - x}{h} \right) \right). \tag{20}$$

Note that, by the usual properties of a kernel of order $[\beta]$, see e.g., [21, Theorem 1.1], for some positive constant C it holds

$$|(16) - \varphi(x)| \leq Ch^\beta.$$

It is assumed that the $\widehat{\theta}_j$'s satisfy (8), thus

$$\begin{aligned}
(17) &= \frac{1}{h} \mathbf{E} \left(\mathbf{E}(\Delta | \theta_1) \frac{1}{h} K' \left(\frac{\theta_1 - x}{h} \right) \right) \\
|(17)| &\leq \frac{C \log n}{nh} \int \frac{1}{h} \left| K' \left(\frac{u - x}{h} \right) \right| \varphi(u) du \\
&\leq \frac{C \log n}{nh} \int |K'(v)| \varphi(x + vh) dv \leq \frac{C \log n}{nh} \|\varphi\|_\infty \int |K'|.
\end{aligned}$$

Splitting (18) using \mathcal{A}_1 and its complement,

$$|(18)| \leq \frac{C \log n}{nh^2} \|\varphi\|_\infty \int |K''|.$$

By the same argument,

$$|(19)| \leq C \sum_{p=3}^{k-1} \left(\sqrt{\frac{\log n}{nh^2}} \right)^p \leq \frac{C \log n}{nh^2},$$

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

$$|(20)| \leq \frac{C}{h} \left(\sqrt{\frac{\log n}{nh^2}} \right)^k \leq \frac{C}{\sqrt{nh}} \left(\frac{\log n}{n^{k-1} h^{2k+1}} \right)^{1/2}$$

Thus

$$b(x)^2 \leq c \left[h^{2\beta} + \frac{1}{nh} \frac{\log^2 n}{nh^3} + \frac{1}{nh} \frac{\log^k n}{n^{k-1} h^{2k+1}} \right]. \quad (21)$$

The variance term is bounded by

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

Finally, choosing k large enough in (21), we obtain, for any x in Θ ,

$$\mathbf{E} ([\widehat{\varphi}(x) - \varphi(x)]^2) \leq c \left[h^{2\beta} + \frac{1}{nh} \frac{\log^2 n}{nh^3} \right] + \frac{C}{J_n h} \|\varphi\|_\infty \|K\|_\infty \quad (22)$$

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

- If $J_n \leq (n/\log n)^{\frac{2\beta+1}{\beta+2}}$, then choosing $h_n = n^{-\frac{1}{2\beta+1}}$ implies that $\frac{1}{nh} \frac{\log^2 n}{nh^3} \leq \frac{C}{J_n h}$, leading to the rate

$$\mathbf{E}([\hat{\varphi}(x) - \varphi(x)]^2) \leq cJ_n^{-\frac{2\beta}{2\beta+1}}.$$

- If $J_n \geq (n/\log n)^{\frac{2\beta+1}{\beta+2}}$, then choosing $h_n = (\log n/n)^{\frac{1}{\beta+2}}$, implies that $\frac{1}{nh} \frac{\log^2 n}{nh^3} \geq \frac{C}{J_n h}$, leading to the rate

$$\mathbf{E}([\hat{\varphi}(x) - \varphi(x)]^2) \leq cn^{-\frac{2\beta}{\beta+2}}.$$

Other choices of h_n can easily be seen to lead to slower rates when optimizing (22). □

Remark 3.4. 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 (3). 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 3.5. Theorem 3.3 requires the conditions $\beta > 1$ and (8) 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 $|(17)| \leq c/(\sqrt{nh})$, 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 [15]. Second, we construct the nonparametric estimator of the density defined in Section 3.2 and illustrate its behavior on both simulated data and real data. We point out that in the considered examples, we deal with the case where $f^{[j]} = f$ which is often used in practice where individual effects is only expressed through a warping effect of a main behaviour modeled by a (common) unknown function f .

4.1 Numerical algorithm for shift estimation and extensions

To compute explicitly $\hat{\theta}_n$ given by (3) for each curve, one has to choose an appropriate filter (h_k) . According to Lemma 2.4, a good choice of weights should make the remainder term $R_n[h, f^{[j]}]$ small. The authors in [7] determined a sequence (h_k) - roughly, a well-chosen sequence of Pinsker weights - such that the second-order term is optimal from the minimax point of view. However, this choice depends on the regularity parameter of the function $f^{[j]}$, which are not known in practice.

Here we use an adaptation of the penalization technique introduced in [15] to determine an appropriate sequence (h_k) . However, contrary to that paper, where only projection weights $h_k = \mathbf{1}_{|k| \leq K}$ are considered, note that the criterion (3) enables the use of a much broader variety of weights (h_k) , provided these satisfy conditions (C)-(T). As explained below, the use of Pinsker weights, see (23), enables a smoothing in the criterion which

improves estimation with respect to [15].

We also would like to mention an alternative method based on a data-driven choice of the filter, proposed in [8]. This method is very interesting in particular from the theoretical point of view, since it achieves an optimal minimax second-order term and is adaptive to the regularity of $f^{[j]}$. But the method is asymptotic in nature and, though it performs well for not too complicated signals, our method seems more appropriate for complicated signals (that is, with possibly many non-zero Fourier coefficients) $f^{[j]}$, as in the laser vibrometry example below.

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

$$h_k = \left[1 - (k/K)^\beta\right]_+, \quad k \geq 0, \quad (23)$$

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

$$\Lambda_K(\tau) = \sum_{k=1}^K h_k \left| \frac{1}{n} \sum_{i=1}^n \cos(2\pi k(t_i - \tau)) Y_i \right|^2, \quad (24)$$

where Y_i , $i = 1, \dots, n$ 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 $1/m$: $\{\tau_1, \dots, \tau_{i+1} = \tau_1 + i/m, \dots, \tau_{max}\}$, 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

$$\begin{aligned} \hat{\tau}(K) &= \operatorname{argmax}_{\tau_1, \dots, \tau_{max}} \Lambda_K(\tau), \\ M(K) &= \operatorname{argmax}_{\tau_1, \dots, \tau_{max}} \Lambda_K(\tau). \end{aligned}$$

Penalization. We would like to find an adapted sequence of weights, or equivalently an integer K . This is done, as in [15] or [4], using a penalization method. Let

$$\hat{K}(\alpha) = \operatorname{argmax}_{K_1, \dots, K_{max}} \{-M(K) + \alpha K\}. \quad (25)$$

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 $K \rightarrow -M(K)$. Let us recall the following lemma from [15].

Lemma 4.1. *There exist two sequences $K_1 = 1 < K_2 < \dots$, and $\alpha_0 = +\infty > \alpha_1 > \dots$, with:*

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

and such that

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

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

$$\tau^* = \operatorname{argmax}_{\tau_i} \sum_{p: \hat{\tau}(K_p)=\tau_i} \{\alpha_{p-1} - \alpha_p\}. \quad (26)$$

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 $\tau^* = \hat{\tau}(\hat{K}(\alpha^*))$.

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 (3). In fact, for k much larger than 100, the corresponding elements in the sum of squares in (3) are mainly noise. As we see in Figure 1, with the choice of $\hat{\tau}$ given in (26), our algorithm chooses \hat{K} in the appropriate interval.

Compared to the method used in [15], the use of the Pinsker-type weights (23) allows a *smoothing* with respect to projection weights, making the detection of main jumps in the convex hull of $K \rightarrow -M(K)$ less sensible to local irregularities of $K \rightarrow M(K)$, which slightly improves estimation. An extensive numerical comparison of the use of the two type of weights, which is beyond the scope of this paper, is carried out in [3] in the case of the period model for discrete design and gains of 10 to 20 % in the estimation of θ are observed for a laser vibrometry example with the unknown f similar to the function f_1 above.

The period model. We note that this algorithm can also be implemented for the period model (9) 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 (24) by its equivalent $\cos(2\pi k(t_i/\tau))$. A numerical study is carried out in [3], leading to similar conclusions than the one presented here.

4.2 Numerical algorithm for density estimation

Once obtained the estimators of the realizations $\hat{\theta}_j$, $j = 1, \dots, J$, we can build the estimator of the density φ defined by (14). 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.

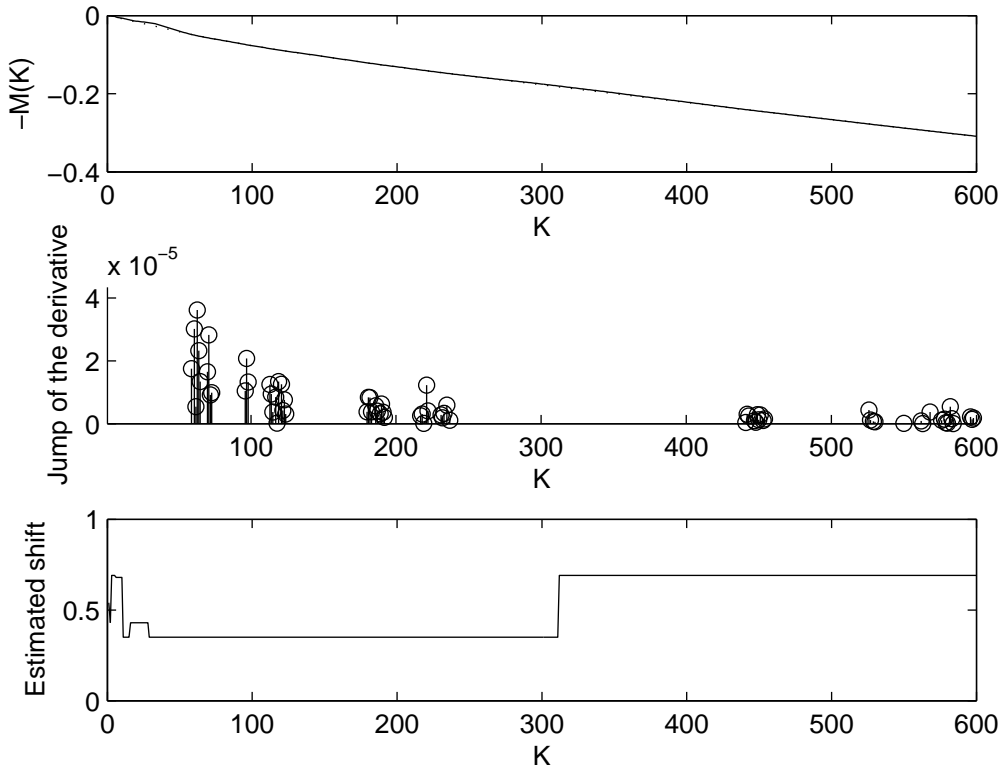


Figure 1: Finding the parameter $\hat{\alpha}$

Simulated data (I). The function f is the sine function on an half period, while the law of the shift θ is a bimodal mixture of compactly supported densities. We perform 50 random translation of the original curve with $n = 100$ observations per curve. In Figure 2, we present the observed curves in model (1). To study the performance of the estimator described in this paper, we first computed the preliminary estimates $\hat{\theta}_j$ obtained by the semiparametric method of Section 2, using the practical algorithm of Section 4. This set of values was then used to build two nonparametric estimators of the density φ , denoted respectively *SPGaussian* and *SPepanech*, using (14) 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.

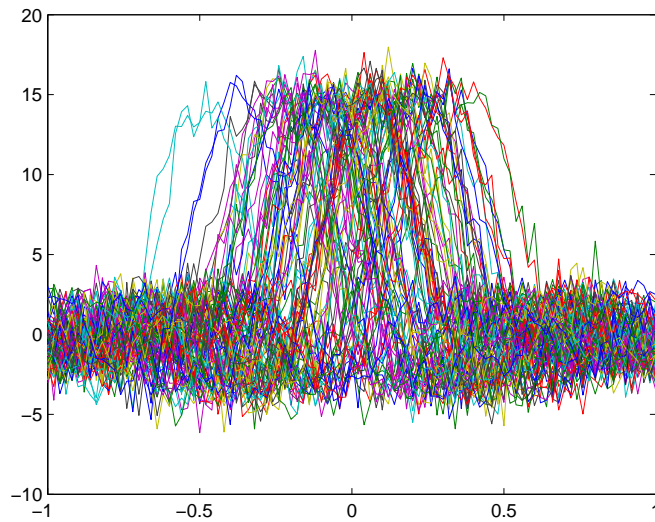


Figure 2: Simulated shifted curves.

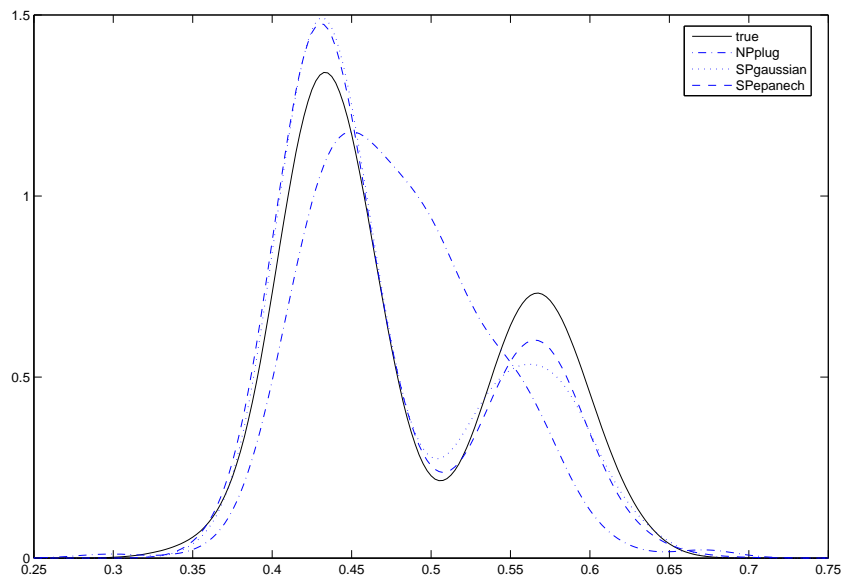


Figure 3: Estimators of the shifts density.

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 [15] 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 (left picture) and both the true shift density φ and the

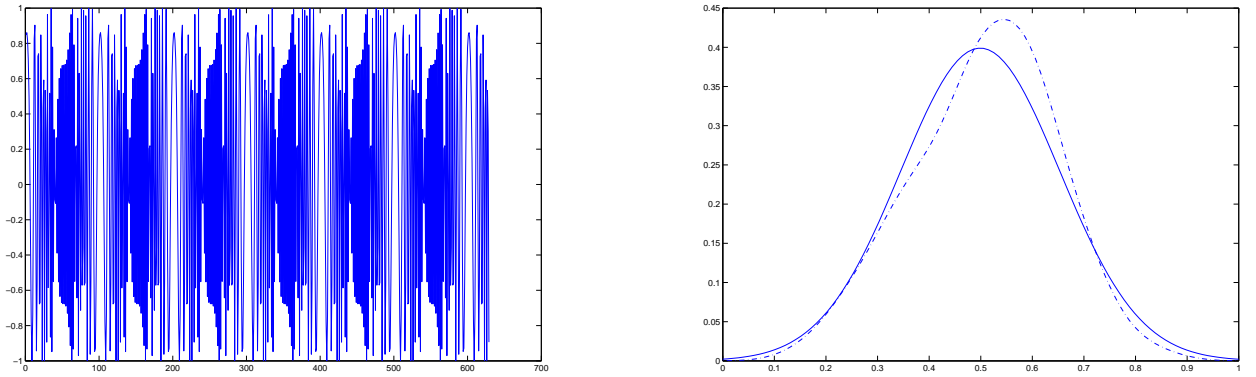


Figure 4: Simulated laser vibrometry-type functions

estimated density, plotted in dotted line (right picture). The curves have been shifted using a compactly supported smooth density φ , with $n = 100$ observations and $J_n = 30$. The two functions, the true function φ and the estimate, 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 velocities of vehicles on a motorway on the suburbs of Paris.

After a preliminary 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 [17]. 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, see the left-hand side of Figure 5. Understanding roadtrafficking behaviour, 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, see the right-hand side of Figure 5, can be later understood as the consequence of different weather conditions on the road network.

5 Appendix

First let us state a useful result about the control of Fourier coefficients by discrete approximations, which will be used in the sequel to control remainder terms. Let us denote

$$\hat{f}_k = \frac{1}{n} \sum_{i=1}^n \sqrt{2} \cos(2\pi k(t_i - \theta)) f(t_i - \theta) \quad (27)$$

$$\hat{g}_k = \frac{1}{n} \sum_{i=1}^n \sqrt{2} \sin(2\pi k(t_i - \theta)) f(t_i - \theta). \quad (28)$$

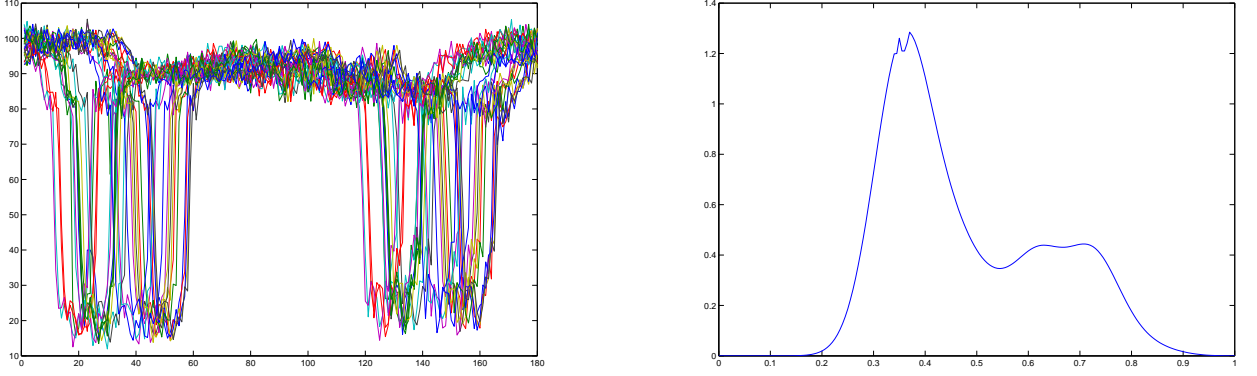


Figure 5: Real data: velocities curves

Lemma 5.1. For any f in the class $F = F(\rho, C_0)$ satisfying (A1)-(A3), there exists a constant C depending only on C_0 such that, for any $k \geq 1$,

$$|\widehat{f}_k - f_k| \leq C \left(\frac{k}{n} \wedge 1 \right) \quad \text{and} \quad |\widehat{g}_k| \leq C \left(\frac{k}{n} \wedge 1 \right), \quad (29)$$

where $a \wedge b$ denotes the minimum of the two reals a and b .

Proof. For any continuously differentiable function φ on the interval $[0, 1]$ and $t_i = i/n$, it holds

$$\left| \frac{1}{n} \sum_{i=1}^n \varphi(t_i) - \int_0^1 \varphi(u) du \right| = \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (\varphi(t_i) - \varphi(u)) du \right|.$$

If $\|\varphi'\|_\infty$ denotes the supremum norm of the derivative of φ on $[0, 1]$, we have

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n \varphi(t_i) - \int_0^1 \varphi(u) du \right| &\leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \|\varphi'\|_\infty |t_i - u| du \\ &\leq \|\varphi'\|_\infty \sum_{i=1}^n (t_i - t_{i-1})^2 / 2 = \frac{\|\varphi'\|_\infty}{2n}. \end{aligned}$$

Now let us apply the preceding to the functions $\varphi_1(u) = \cos(2\pi k(u - \theta))f(u - \theta)$ and $\varphi_2(u) = \sin(2\pi k(u - \theta))f(u - \theta)$ respectively. By symmetry and 1-periodicity of f , we have $\int_0^1 \varphi_2(u) du = 0$. For any real u ,

$$|\varphi_1'(u)| \leq 2\pi k \|f\|_\infty + \|f'\|_\infty,$$

and, similarly, the same bound holds for $|\varphi_2'(u)|$. Now observe that $\|f\|_\infty$ and $\|f'\|_\infty$ are bounded if f belongs to the class F . Indeed, if $f \in F$, then f' is continuously differentiable and 1-periodic thus it is the limit of its Fourier series. For any $u \in [0, 1]$, we have

$$f'(u) = \sum_{k \geq 1} (-2\pi k) f_k \sin(2\pi k u).$$

To see that the latter quantity is bounded it suffices to check that the series $\sum k f_k$ converges. This is a consequence of Cauchy-Schwarz inequality and (A3) since

$$\sum_{k \geq 1} k |f_k| \leq \left(\sum_{k \geq 1} k^4 f_k^2 \right)^{1/2} \left(\sum_{k \geq 1} k^{-2} \right)^{1/2}$$

is bounded. Similarly, $\|f\|_\infty$ is bounded by $\sum_{k \geq 1} |f_k|$, which is bounded if f is in F , which implies that $\widehat{f}_k - f_k$ and \widehat{g}_k are bounded by a constant times k/n . The fact that they are also bounded by a constant follows from the fact that $\|f\|_\infty$ is bounded over F . \square

Lemma 5.2. *Assume that conditions (A1)-(A3) and (C1)-(C3) are fulfilled. Then for some positive constant C , denoting $\|h'\|^2 = \sum_{k \geq 1} (2\pi k)^2 h_k^2$,*

$$\sum_{k \geq 1} h_k^2 k^6 \leq Cn^2, \quad \sum_{k \geq 1} h_k k^2 |f_k| \leq C \frac{\|h'\|}{\log^2 n} \quad \text{and} \quad \sum_{k \geq 1} h_k k^2 |f_k (f_k - \widehat{f}_k)| \leq C \frac{\|h'\|}{n}.$$

Proof of Lemma 5.2. Note that for any integer k , we have $h_k^2 k^6 \leq h_k k^4 \max_{k \geq 1} h_k k^2$. The latter maximum is smaller than the corresponding sum over k which, due to (C3), is at most $D_1 n$. Using (C3) again, one obtains the first inequality. Then due to (C2),

$$\sum_{k \geq 1} h_k k^2 |f_k| \leq \left(\max_{k \geq 1} h_k k \right) \sum_{k \geq 1} k |f_k| \leq C \frac{\|h'\|}{\log^2 n},$$

which yields the second inequality. Finally, using (29) and the Cauchy-Schwarz inequality,

$$\sum_{k \geq 1} h_k k^2 |f_k (f_k - \widehat{f}_k)| \leq \frac{C}{n} \sum_{k \geq 1} h_k k^3 |f_k| \leq \frac{C}{n} \left(\sum_{k \geq 1} h_k^2 k^2 \right)^{1/2} \left(\sum_{k \geq 1} k^2 f_k^4 \right)^{1/2},$$

which yields the third inequality using (A3). \square

We can now turn to the proofs of Lemmas 2.3, 2.4 and 2.5. We shall work conditionally to the event $\theta_j = \theta$. To abbreviate the notation, we omit the index j and drop the notation $|\theta_j$). Thus the expectations in the following should be understood at fixed θ_j .

The main novelty with respect to [7] consists in proving that the arguments used by the authors in that paper can be adapted in the discrete setting, by showing that the arguments still go through when working with the discrete approximations \widehat{f}_k and \widehat{g}_k instead of f_k and 0 respectively.

Proof of Lemma 2.3. The contrast function to maximize is, according to (3),

$$\begin{aligned} L(\tau) &= \sum_{k \geq 1} h_k \left[\frac{1}{n} \sum_{i=1}^n \sqrt{2} \cos(2\pi k(t_i - \tau)) Y_i \right]^2 \\ &= \sum_{k \geq 1} h_k \left[\cos(2\pi k(\theta - \tau)) \widehat{f}_k - \sin(2\pi k(\theta - \tau)) \widehat{g}_k \right. \\ &\quad \left. + \frac{1}{\sqrt{n}} (\cos(2\pi k\tau) \xi_k + \sin(2\pi k\tau) \xi_k^*) \right]^2, \end{aligned} \quad (30)$$

where \widehat{f}_k and \widehat{g}_k are defined by (27)-(28). The criterion $L(\tau)$ 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

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

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

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

Note that this is the analog of the decomposition of [7], p. 185, except that here the quantity $\cos(2\pi k(\theta - \tau)) \widehat{f}_k + \sin(2\pi k(\theta - \tau)) \widehat{g}_k$ replaces $\cos(2\pi k(\theta - \tau)) f_k$. Let us see how the argument is further modified.

The stochastic term η_2 is exactly the same as in [7]. The term η_1 is such that its derivative η_1' is a zero-mean stationary Gaussian process and one has

$$\mathbf{E}(\eta_1'(\tau)^2) = \|f'\|^{-2} \sum_{k \geq 1} h_k^2 (2\pi k)^2 (\widehat{f}_k^2 + \widehat{g}_k^2),$$

$$\mathbf{E}(\eta_1''(\tau)^2) = 4\|f'\|^{-2} \sum_{k \geq 1} h_k^2 (2\pi k)^4 (\widehat{f}_k^2 + \widehat{g}_k^2).$$

So, $\eta_1'(\tau)$ has a variance bounded from below by a constant times \widehat{f}_1^2 which is bounded away from zero for n large enough due to (29) and (A3). Moreover, the variance of η_1'' is bounded. Hence one can apply Rice formula as in [7] to obtain that there are some positive constants C and D , such that for all $x > 0$,

$$\mathbf{P}(\sup_{\tau \in \Theta} |\eta_1'(\tau)| > x) \leq C \exp(-Dx^2), \quad (31)$$

which is the result obtained in [7]. Finally we deal with η_0 by writing

$$\eta_0(\tau) = \sum_{k \geq 1} h_k \left[\cos^2(2\pi k(\tau - \theta)) \widehat{f}_k^2 - 2 \cos(2\pi k(\tau - \theta)) \sin(2\pi k(\tau - \theta)) \widehat{f}_k \widehat{g}_k + \sin^2(2\pi k(\tau - \theta)) \widehat{g}_k^2 \right] = \gamma_0(\tau) + \gamma_1(\tau) + \gamma_2(\tau).$$

We have that $\gamma_0'(\theta) = 0$ and $\gamma_0''(\theta) \leq -(2\pi)^2 \widehat{f}_1^2$ is bounded away from zero due to (29) and (A3). Thus, similarly to [7], one has $\gamma_0(\tau) - \gamma_0(\theta) \leq -C|\tau - \theta|^2$ for all $\tau \in \Theta$. Now note that for all real τ ,

$$|\gamma_1(\tau) + \gamma_2(\tau)| \leq \sum_{k \geq 1} 2\widehat{f}_k \widehat{g}_k + \widehat{g}_k^2.$$

Hence using (29), the sum $\gamma_1(\tau) + \gamma_2(\tau)$ is a $O(1/n)$ uniformly in τ . The argument is now completed as follows. Using the obtained bounds, for any positive x ,

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

Now setting $x = x_n = K(\log n)^{1/2}$, for some positive constants C_1 and C_2 we have

$$\begin{aligned} & \mathbf{P}_\theta \left(|\hat{\theta} - \theta| \sqrt{n \|f'\|^2} > x_n \right) \\ & \leq \mathbf{P}_\theta \left(\frac{\sqrt{n} \|f'\|}{x_n} \sup_{t \in \Theta} \left\{ 2 \frac{\|f'\| |\eta'_1(t)|}{\sqrt{n}} + \frac{|\eta'_2(t)|}{n} \right\} \geq -C + \frac{O(n^{-1})}{x_n} \right) \\ & \leq \mathbf{P}_\theta \left(\sup_{t \in \Theta} |\eta'_1(t)| \geq C x_n \right) + \mathbf{P}_\theta \left(\sup_{t \in \Theta} |\eta'_2(t)| \geq C \sqrt{n} x_n \right). \end{aligned}$$

This is further bounded using (31) for the first term and Lemma 3 in [7] for the second term, which concludes the proof. \square

Before we turn to the proof of Lemma 2.4, we state a Lemma which summarizes the properties of the criterion L , see Equation (30), and its derivatives. Note that it is in particular a natural adaptation of Lemma 6 in [7].

Lemma 5.3. *Uniformly over $\theta \in \Theta$ and $f \in F$, as $n \rightarrow +\infty$,*

$$\mathbf{E}(L'(\theta)^2) = \frac{4}{n} \left(\sum_{k \geq 1} (2\pi k)^2 h_k^2 f_k^2 + (1 + o(1)) \frac{\|h'\|^2}{n} \right) \quad (32)$$

$$\mathbf{E}(L''(\theta)) = -2 \sum_{k \geq 1} h_k (2\pi k)^2 f_k^2 + o\left(\frac{\|h'\|^2}{n}\right). \quad (33)$$

Moreover, uniformly over $\theta \in \Theta$ and $f \in F$, as $n \rightarrow +\infty$,

$$\mathbf{E}\{L''(\theta) - \mathbf{E}(L''(\theta))\}^2 = O(n^{-1}) \quad \text{and} \quad \mathbf{E} \left(\sup_{\zeta \in \Theta} L^{(3)}(\zeta)^2 \right) = O(1). \quad (34)$$

Proof of Lemma 5.3. Let us denote

$$\xi_k(\theta) = \frac{1}{n} \sum_{i=1}^n \sqrt{2} \cos(2\pi k(t_i - \theta)) \varepsilon_i \quad \text{and} \quad \xi_k^*(\theta) = \frac{1}{n} \sum_{i=1}^n \sqrt{2} \sin(2\pi k(t_i - \theta)) \varepsilon_i.$$

Simple calculations from (30) lead to

$$L'(\theta) = 2 \sum_{k \geq 1} h_k (2\pi k) (\hat{f}_k + n^{-1/2} \xi_k(\theta)) (\hat{g}_k + n^{-1/2} \xi_k^*(\theta)) \quad (35)$$

$$L''(\theta) = 2 \sum_{k \geq 1} h_k (2\pi k)^2 \left\{ -(\hat{f}_k + n^{-1/2} \xi_k(\theta))^2 + (\hat{g}_k + n^{-1/2} \xi_k^*(\theta))^2 \right\} \quad (36)$$

From (35) we deduce that

$$\begin{aligned} \mathbf{E}(L'(\theta)^2) &= 4 \sum_{k \geq 1} h_k^2 (2\pi k)^2 \left(\frac{f_k^2}{n} + \frac{1}{n^2} \right) \\ &+ 4 \sum_{k \geq 1} h_k^2 (2\pi k)^2 \left(\hat{f}_k^2 \hat{g}_k^2 + \frac{1}{n} \{ 2f_k (\hat{f}_k - f_k) + (\hat{f}_k - f_k)^2 \} + \frac{\hat{g}_k^2}{n} \right) \\ &+ 8 \sum_{k \neq l} h_k h_l (2\pi k) (2\pi l) \hat{f}_k \hat{f}_l \hat{g}_k \hat{g}_l. \end{aligned}$$

The second term in the last display is bounded using first (29) and then Lemma 5.2, by $O(\|h'\|/n^2)$, which is a $o(\|h'\|^2/n^2)$ since due to (C2), the norm $\|h'\|$ tends to $+\infty$. To bound the third term, note that it is bounded above by the square of

$$\sum_{k \geq 1} h_k k |\widehat{f}_k \widehat{g}_k| \leq \frac{C}{n} \sum_{k \geq 1} h_k \left(k^2 |f_k| + \frac{k^3}{n} \right) \leq \frac{C}{n} \left(\frac{\|h'\|}{\log^2 n} + 1 \right),$$

using Lemma 5.2. The corresponding square is thus a $o(\|h'\|^2/n^2)$. For the second derivative, from (36) we deduce that

$$\begin{aligned} \mathbf{E}(L''(\theta)) &= 2 \sum_{k \geq 1} h_k (2\pi k)^2 (-\widehat{f}_k^2 + \widehat{g}_k^2) \\ \mathbf{E}\{L''(\theta) - \mathbf{E}(L''(\theta))\}^2 &= \frac{16}{n} \sum_{k \geq 1} h_k^2 (2\pi k)^4 (\widehat{f}_k^2 + \widehat{g}_k^2 + \frac{1}{n}) \end{aligned} \quad (37)$$

and proceeding as for (35), using (29) and Lemma 5.2, one obtains (33) and the first part of (34). Finally, the result about $L^{(3)}$ is obtained as follows. Proceeding as in Lemma 6 in [7], one easily sees that

$$\sup_{\zeta \in \Theta} |L^{(3)}(\zeta)| \leq C \sum_{k \geq 1} h_k k^3 \left(\widehat{f}_k^2 + \widehat{g}_k^2 + \frac{1}{n} (\xi_k^2 + \xi_k^{*2}) \right).$$

The deterministic part of the last display is bounded by a constant times

$$\sum_{k \geq 1} h_k k^3 \left(2f_k^2 + 2(\widehat{f}_k - f_k)^2 + \widehat{g}_k^2 \right) \leq C \sum_{k \geq 1} h_k \left(k^3 f_k^2 + k^3 \frac{k}{n} \right) \leq C'.$$

To obtain the first inequality, using (29) we have bounded one $\widehat{f}_k - f_k$ and one \widehat{g}_k by Ck/n and the other ones by a constant. The second inequality is obtained using (A3) and (C3), which concludes the proof of the Lemma. \square

Note that all the dependence in \widehat{f}_k and \widehat{g}_k has vanished in Lemma 5.3, replaced by results in function of f_k only. In fact, the results of this Lemma are exactly the ones used in [7] to prove the second order expansion, so in fact using this observation there is nothing left to prove to obtain Lemma 2.4. We include however the end of the proof for completeness.

Proof of Lemma 2.4. As in [7], the proof is in two steps. The first step is to prove that $\widehat{\tau}$ defined by the following relation has the desired second order expansion,

$$L'(\theta) + (\widehat{\tau} - \theta) \mathbf{E}(L''(\theta)) = 0. \quad (38)$$

Let us evaluate $\mathbf{E}((\widehat{\tau} - \theta)^2 I_n(f))$, where $I_n(f) = n \|f'\|^2$. From (32) and (33) it follows

$$\begin{aligned} \mathbf{E}((\widehat{\tau} - \theta)^2 I_n(f)) &= \left[1 + \|f'\|^{-2} \sum_{k \geq 1} (h_k^2 - 1) (2\pi k)^2 f_k^2 + \|f'\|^{-2} \|h'\|^2 (1 + o(1))/n \right] \\ &\quad \times \left[1 + \|f'\|^{-2} \sum_{k \geq 1} (h_k - 1) (2\pi k)^2 f_k^2 + o(\|h'\|^2/n) \right]^{-2}. \end{aligned}$$

Let us expand the square at the denominator and then use a Taylor expansion of the function $x \rightarrow (1+x)^{-1}$. Then, one computes the product with the numerator and uses assumption (T). One obtains

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

which is the desired expansion for $\hat{\tau}$.

In a second step, we prove that $\hat{\theta}$ and $\hat{\tau}$ are close enough. It is sufficient to do this on the set $\mathcal{A}_1 = \{|\hat{\theta} - \theta| \leq D(n^{-1} \log n)^{1/2}\}$ since the probability of its complement is negligible due to Lemma 2.3. By definition of $\hat{\theta}$, we have $L'(\hat{\theta}) = 0$. By Taylor's expansion,

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

for some random variable ζ , which can also be written

$$\begin{aligned} 0 &= L'(\theta) + (\hat{\theta} - \theta)\mathbf{E}(L''(\theta)) \\ &\quad + (\hat{\theta} - \theta)[L''(\theta) - \mathbf{E}(L''(\theta))] + \frac{(\hat{\theta} - \theta)^2}{2} L^{(3)}(\zeta). \end{aligned} \quad (39)$$

Subtracting (38) and (39), one obtains

$$\begin{aligned} &\mathbf{E}\left((\hat{\theta} - \hat{\tau})^2 \mathbf{1}_{\mathcal{A}_1}\right) \mathbf{E}(L''(\theta))^2 \\ &\leq 2\mathbf{E}\left((\hat{\theta} - \theta)^2 \{L''(\theta) - \mathbf{E}(L''(\theta))\}^2 \mathbf{1}_{\mathcal{A}_1}\right) + \mathbf{E}\left((\hat{\theta} - \theta)^4 \sup_{\zeta \in \Theta} |L^{(3)}(\zeta)|^2 \mathbf{1}_{\mathcal{A}_1}\right). \end{aligned}$$

Using (34) and the definition of \mathcal{A}_1 , one obtains $\mathbf{E}((\hat{\theta} - \hat{\tau})^2 I_n(f) \mathbf{1}_{\mathcal{A}_1}) \leq Cn^{-1} \log^2 n$ which is a $o(R_n(h, f))$. Finally, by similar arguments, one also sees that $\mathbf{E}((\hat{\theta} - \hat{\tau})(\hat{\tau} - \theta) \mathbf{1}_{\mathcal{A}_1})$ is a $o(R_n(h, f))$ which concludes the proof. \square

Proof of Lemma 2.5:

Proof. Starting from (39), using the triangle and Cauchy-Schwarz inequalities,

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

The first term on the right-hand side of the last display can be bounded using (29),

$$\mathbf{E}(L'(\theta)) = \left| 2 \sum_{k \geq 1} h_k(2\pi k) \hat{f}_k \hat{g}_k \right| \leq C \sum_{k \geq 1} h_k \left(k^2 \frac{|f_k|}{n} + \frac{k^3}{n^2} \right),$$

which is a $O(n^{-1})$ due to the assumption on f and (C3). To bound the second term, we use (34) and the fact that the result of Lemma 2.4 implies

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

We note that the latter relation could also be checked directly, similarly to the proof of Lemma 2.4 but without keeping second order terms. To bound the third term, we use Cauchy-Schwarz inequality

$$\mathbf{E} \left\{ (\widehat{\theta} - \theta)^2 \sup_{\zeta \in \Theta} |L^{(3)}(\zeta)| \right\} \leq \mathbf{E} \left((\widehat{\theta} - \theta)^4 \right)^{1/2} \mathbf{E} \left\{ \sup_{\zeta \in \Theta} |L^{(3)}(\zeta)|^2 \right\}^{1/2}.$$

Now with $\mathcal{A}_1 = \{|\widehat{\theta} - \theta| \leq D(n^{-1} \log n)^{1/2}\}$, due to Lemma 2.3,

$$\begin{aligned} \mathbf{E}((\widehat{\theta} - \theta)^4) &= \mathbf{E}((\widehat{\theta} - \theta)^4 \mathbf{1}_{\mathcal{A}_1}) + \mathbf{E}((\widehat{\theta} - \theta)^4 \mathbf{1}_{\mathcal{A}_1^c}) \\ &\leq C(\log n/n)^2 + C \exp(-dD^2 \log n). \end{aligned}$$

Choosing D large enough, we obtain that the third term is a $O(\log n/n)$. The fact that $|\mathbf{E}(L''(\theta))|$ is bounded from above and below by positive constants, which follows from (37) and (A3)-(C3), yields the announced result. \square

Acknowledgments. The authors are grateful to Professor Alexandre Tsybakov for an insightful remark concerning this work. We also thank a referee whose comments lead us to correct a mistake in a previous 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 semi-parametrics *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] D. Chafai and J-M. Loubes. Maximum likelihood for a certain class of inverse problems: an application to pharmacokinetics. *SPL*, 76:1225–1237, 2006.
- [7] 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.
- [8] A. S. Dalalyan, Stein shrinkage and second-order efficiency for semiparametric estimation of the shift, *Math. Methods Statist.*, 161, 42–62, 2007.
- [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 Statistics*, 8:387–419, 2003.

- [11] F. Gamboa and J-M. Loubes and E. Maza. Shifts estimation with M-estimators. *Electronic Journal of Statistics*, 616–640, 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] 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.
- [16] B. Lindsay. The geometry of mixture likelihoods: a general theory. *Ann. Statist.*, 11(1):86–94, 1983.
- [17] 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.
- [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. Rønne. 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.
-