Statistica Sinica **19** (2009), 667-684

WAVELET ESTIMATIONS FOR LONGITUDINAL DATA

Trung Tu Nguyen

University of Paris VII

Abstract: In the context of longitudinal data analysis, we study a random function X that represents patients or subjects observed at randomly distributed points. Principal components analysis (PCA) is useful in understanding the random effects of X. In this paper, we estimate the mean and covariance functions of X by wavelet methods. Proposed wavelet estimators give interesting performances over a wide class of functions, even if the regularity parameters of the original functions are not assumed to be greater than 2. In another problem of longitudinal analysis, we study the regression of observations at recorded times, when the number of observations per unit is assumed to be a finite integer. In this context, under Gaussian assumptions, our wavelet estimator could be proved to be optimal.

Key words and phrases: Karhunen-Loève expansion, nonparametric estimation, projection estimation, principal components methods, wavelet decomposition.

1. Introduction

Research in different disciplines involves data from high-dimensional repeated measurements of the same object. Depending on how measurements are recorded across times, densely or sparsely, we have different ways to analyze and process them.

When measurements are taken on a dense time grid, data typically consist of one observation function per each subject. The statistical analysis of a sample of such graphs is called "Functional Data Analysis" (FDA).

On the other hand, when data are recorded at sparse time points, for each subject we have only, say, m observations at m time points. Moreover, data always come with experimental errors. Longitudinal data arise commonly in health sciences and engineering applications, under different terminology, with the defining characteristic that individuals are measured repeatedly through time. Longitudinal data analysis (LDA) is similar to FDA in that one can suppose an underlying function exists. The difference between their analyses comes out in two ways: the errors, and the continuity of the observations. In many different models, these points are more conceptual than actual. However, longitudinal data require special statistical methods because the observations of an individual are correlated, and this must be taken into account.

1.1 Considered models

In this paper, LDA is scrutinized with the model

$$Y_{ij} = X_i(T_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, n \quad j = 1, \ldots, m,$$

where the observations Y_{ij} of n subjects are recorded at random times T_{ij} with experimental errors ϵ_{ij}. For each subject i, we have m data points (m is usually small). Our aim is to estimate the behaviour of the random process X from the nxm-matrix of observations $(Y_{ij}, T_{ij})_{i,j}$. In this context, principal components analysis (PCA) is very useful in explaining the effects of discretely recorded results on statistical estimators. PCA is a technique for simplifying a dataset, by reducing it to a few essential dimensions. It was studied in early works by Grenander (1968) and more recently by Rice and Silverman (1991), Ramsay and Silverman (1997), and many others. Nonparametric methods for unbalanced longitudinal data were studied by Boularan, Ferré and Vieu (1995).

Yao, Müller and Wang (2005) proposed a procedure to estimate functional principal component scores for sparsely recorded data. They used the local weighted polynomial smoothing method to estimate the mean function $\mu(u) = E\,(X(u))$, and the fitting of local lines (planes) (Fan and Gijbels (1996)) to estimate the covariance function $\omega(u, v) = \mathrm{Cov}\;(X(u), X(v))$. These techniques are applicable for FDA problems but, in the case of LDA, some special conditions on the random process X and its derivatives are required.

In this paper, we introduce wavelet estimators for the mean and the covariance, and we investigate their performances under L_2 risk. Without any conditions on the derivatives of the process X, our estimators give interesting performances.

In mean estimation, an approach using a functional regression model is chosen. Under a few conditions on the process X, our wavelet estimator $\hat{\mu}$ of the mean function μ has the following rates of convergence (over Besov classes):

$$\left[(nm)^{-\frac{2s}{2s+1}} + \frac{\int_0^1 \mathrm{Var}\,(X(u))du}{n}\right],$$

where s is the regularity parameter of the mean function. We remark that, for an extreme case where $n << m \to \infty$, our rates of convergence become $(\int_0^1 \mathrm{Var}\,(X(u))du/n)$, which corresponds to the case where n entire graphs $X_i(u)$ are observed. For another extreme case where only few observations of each individual are observed ($m << n$), especially when m is a finite integer, the rates of convergence would be proved to be minimax for the uniform design under Gaussian assumptions (see more details in Comment 3.2).

For the covariance estimations, we have two results.

- If the mean function is known, then our wavelet estimator gives rates of convergence of $(nm)^{-\gamma/(\gamma+1)}$, where γ is the regularity parameter of the covariance function.
- If the mean function is unknown, then two unknown functions μ and ω need to be estimated consecutively. In this case, we get more complicated rates of convergence that depend on both regularity parameters, s of μ and γ of ω (see more details in Corollary 4.3 and Comments 4.7).

1.2. Discussion

In this subsection, we explain the choice of linear wavelet methods, and we present a more complete schema of the wavelet estimation for longitudinal data.

- We begin in the general case where the only condition on X is $\sup_{u\in[0,1]}$ Var $X(u) = \|\omega\|_\infty < \infty$; this is simple, and easily verified in many applied problem. Assuming more conditions on X could help us deal with technical difficulties (for example, in Subsection 3.3, we use Gaussian properties to simplify the Kullback divergence, which plays the leading role in lower bound controls) but, on the other hand, we lose the generality of the problem.
- We have observations Y_{ij} from n independent individuals. For each individual i, m observations are taken at m random recorded time points. These m observations are correlated, which makes the problem more interesting and also more complicated. Analyzing these correlations is also one of main goals of LDA. However, these correlations also disturb analysis of X's behavior, especially when the number of correlated observations goes to infinity. A simple method is used here to obtain good estimates of X's behavior even as m goes to infinity. We use the projection method to estimate the mean and covariance functions. Upper bounds of the quadratic risk can be obtained for any orthogonal basis. Using wavelet bases could simplify the upper bounds, as shown in the covariance estimations. In addition, wavelet estimations give us the idea of applying wavelet shrinkage procedures, which has the following benefits,
  1. Building adaptive estimators-adaptive procedures which do not depend on the regularity parameters of mean and covariance functions.
  2. Instead of being limited to quadratic risk, we can investigate the performance of wavelet estimators in L_p, with $p > 1$. Moreover, shrinkage wavelet estimators have better convergence rates over function space $B^s_{p,q}$ where $1 < p < 2$, than linear estimators.

However, wavelet shrinkage procedures pose other difficulties. Especially, the wavelet estimated coefficients require more conditions and properties, concentration inequalities for example, which can be difficult when the number of correlated observations goes to infinity. Even in the case where m is a finite integer, more assumptions on the process X are required to deal with technical difficulties. A more complete wavelet estimation schema for LDA could be decomposed into several steps, with different results and different interpretations for each step.

1. *First step:* we present non-adaptive linear wavelet procedures for mean and covariance estimations. In this step, general conditions apply to the number of correlated observations, m, and to the random process X. For the extreme cases, $m = 1$ and $m = \infty$, the convergence rates of the mean estimation can be obtained. Moreover, a minimax result of mean estimation is introduced and proved for the Gaussian case with a finite integer m. We could have a more complicated result for the mixed model, where the mean and covariance functions are unknown and must be consecutively estimated.
2. *Second step:* Adaptive results for mean and covariance estimations (known-mean case), using wavelet shrinkage procedures can be obtained under the following assumptions:
   - m is a finite integer.
   - X is a bounded random process or a Gaussian process with moment conditions.
   - *Brownian motion:* When m is not a finite integer, more information about X could be required to control the behavior of wavelet estimated coefficients; an interesting adaptive result for mean estimation can be proved when the zero-mean random process $\delta(u) = X(u) - E(X(u))$ is a Brownian motion.

This paper presents the first step of this wavelet estimation schema. Results and difficulties from the second step are to be presented and detailed in our next papers. The remainder of this paper is organized as follows. In Section 2, we present briefly the model of longitudinal data and the PCA method. In Section 3, we introduce the wavelet estimator of the mean function and its minimax performance. The wavelet estimator of the covariance function is introduced in Section 4. A short proof of Theorem 3.1 is presented in Section 5.

2. Model and PCA Method

Let $X, X_1, \ldots, X_n$ be independent and identically distributed random functions on $[0, 1]$, satisfying $\int_0^1 \mathrm{Var}\,(X(u))du < \infty$. We consider the model

$$Y_{ij} = X_i(T_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, n, j = 1, \ldots, m, m \geq 2, \tag{2.1}$$

where the observations Y_{ij} are recorded at time points T_{ij} with zero-mean errors ϵ_{ij}. Usually, the noises ϵ_{ij} are assumed to have normal distribution; here the Gaussian assumption is not required except in Section 3.3. We assume that (see also Yao, Müller and Wang (2005)) the following hold.

- The T_{ij} are independent and uniformly distributed in $[0, 1]$.
- The ϵ_{ij} are i.i.d. variables with zero mean and finite variance $E(\epsilon_{11}^2) = \sigma^2 < \infty$.
- The X_i, T_{ij}, and ϵ_{ij} are independent.

A main objective of longitudinal data analysis is to study the behaviour of the random function X. PCA is a useful technique to explain the random effects of X, which are briefly presented in the next subsection.

2.1. PCA method

The PCA method is based on interpreting the covariance function $\omega(u, v) = cov\,(X(u), X(v))$ as the kernel of a linear mapping on the space $L_2([0, 1])$ of square-integrable functions on $[0, 1]$, taking $\alpha(.)$ to $\omega\alpha(.)$ defined by $(\omega\alpha)(u) = \int_{\mathcal{I}} \alpha(v)\omega(u, v)dv$.

Using Mercer's theorem (e.g., Indritz (1963, Chap. 4)), we can write

$$\omega(u, v) = \sum_{j=1}^{\infty} \lambda_j^2 \eta_j(u)\eta_j(v),$$

where λ_j^2 are the positive ordered eigenvalues ($\lambda_1^2 \geq \cdots \geq 0$), and η_j are the corresponding eigenfunctions of the function $\omega(u, v)$.

Because the eigenfunctions η_j form a complete orthonormal sequence of $L_2(\mathcal{I})$, for each i we have

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \zeta_{ij}\eta_j(t), \tag{2.2}$$

where $\mu(t) = E(X_i(t)) = E(X(t))$, and $\zeta_{ij} = \int_0^1 X_i(t)\eta_j(t)dt$ is the jth random effect of the ith subject (for each i, the random variables ζ_{ij} are uncorrelated because of the orthogonality between η_j and η_k,$j \neq k$). This decomposition is also called the "*functional principal component expansion*" and helps us understand the behaviour of X. For example, when the eigenfunction η_1 admits a turning point in a part of $[0, 1]$ where the other eigenfunctions are mostly flat, this turning point is likely to appear with high probability in the random function X. Again, if all eigenfunctions are close to zero in a given part of $[0, 1]$, we may conclude that X is close to its mean in this region.

Unfortunately in the model (2.1), the functions μ and ω that play the leading role in the PCA method are unknown; we must estimate these functions from the data (Y_{ij}, T_{ij}). In the present paper, we investigate the performance of projection estimators under L_2 risk. Section 3 presents the projection estimator of the mean function μ and its asymptotic performance when a wavelet basis is used. The covariance estimation is presented in Section 4.

3. Estimation of the Mean Function

Here we estimate the mean function μ. The projection estimator and its asymptotic behaviour are given in Subsection 3.1 (Thm. 3.1). In Subsection 3.2, we apply the projection procedure for a compactly supported wavelet basis; in that case, interesting rates of convergence can be proved for all bounded μ in a Besov ball (Corollary 3.1).

3.1 Projection estimation for the longitudinal data model

The main idea is to project the observations Y on a basis where the signal f is well concentrated. If $\mathcal{E} = (e_k)_{k\in\mathbb{N}}$ is an orthonormal basis of $L_2([0, 1])$, and $\mu \in L_2([0, 1])$ we can write

$$\mu(x) = \sum_{k\in\mathbb{N}} \alpha_k e_k(x) \quad \textit{with} \quad \alpha_k = \int_0^1 \mu(x)e_k(x)dx.$$

For all finite integers $M > 0$, take

$$P_M(\mathcal{E})\mu(x) = \sum_{k=1}^{M} \alpha_k e_k(x).$$

The idea of this projection estimation is simple: approximate μ by its projection onto the first M functions of the basis $\mathcal{E}$, and replace the coefficients $(\alpha_k)_{k\le M}$ by their estimators $(\hat{\alpha}_k)_{k\le M}$.

From the longitudinal data (Y_{il}, T_{il}), we build the coefficient estimators

$$\hat{\alpha}_k = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij}e_k(T_{ij}), \quad k = 1, \ldots, m.$$

For each individual i, using the data (Y_{il}, T_{il}) (i fixed, j=1,. . . ,m), we have the auxiliary coefficient estimators

$$\hat{\alpha}_{ki} = \frac{1}{m}\sum_{j=1}^{m} Y_{ij}e_k(T_{ij}), \quad k = 1, \ldots, m.$$

The projection estimator is written as

$$\hat{\mu}_M(\mathcal{E})(x) = \sum_{k=1}^{M} \hat{\alpha}_k e_k(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_{Mi}(\mathcal{E})(x), \tag{3.1}$$

where $\hat{\mu}_{Mi}(\mathcal{E})(x) = \sum_{k=1}^{M} \hat{\alpha}_{ki} e_k(x)$.

For all finite integers $M > 0$, we investigate the asymptotic behaviour of $\hat{\mu}_M$. The choice of M depends on the basis $\mathcal{E}$, and is discussed later.

Theorem 3.1.(Uniform Case) *Suppose that* $\omega(u,u) = \mathrm{Var}\,(X(u)) \le \|\omega\|_\infty < \infty, \forall u \in [0,1]$. *For all bounded* $\mu \in L_2([0,1])$, *there exists a positive constant* C, *not depending on* M, *such that*

$$E\|\hat{\mu}_M(\mathcal{E}) - P_M(\mathcal{E})\mu\|_2^2 \le C\left[\frac{M}{nm}(\|\mu\|_\infty^2 + \sigma^2 + \|\omega\|_\infty) + \frac{2\int_0^1 \mathrm{Var}\;(X(u))\,du}{n}\right]. \tag{3.2}$$

Comment 3.1. According to Theorem 3.1, the distance between the estimator $\hat{\mu}_M(\mathcal{E})$ and the projection $P_M(\mathcal{E})\mu$ admits an upper bound that depends on the number of experimental units (n) and on the number of observations per unit (m), but does not depend on the orthonormal basis $\mathcal{E}$.

We remark that our main interest is in the L_2 risk

$$R_2(\hat{\mu}_m, \mu) = E\|\hat{\mu}_M(\mathcal{E}) - \mu\|_2^2 = \int_0^1 (\hat{\mu}_m(u) - \mu(u))^2 du.$$

Note that $R_2(\hat{\mu}_m, \mu) \le 2(E\|\hat{\mu}_M(\mathcal{E}) - P_M(\mathcal{E})\mu\|_2^2 + E\|P_M(\mathcal{E})\mu - \mu\|_2^2)$, where the first term is bounded using Theorem 3.1. In the next section, we will see that the second term can be bounded if we apply the projection procedure for a reasonable basis. Then the finite integer M will be chosen to equilibrate these two terms. This choice is discussed later.

3.2. Performance of wavelet estimation on a Besov class-upper bound

In this section, we illustrate the previous theory with the case where $\mathcal{E}$ is a compactly supported wavelet basis on $[0,1]$, and we consider the performance of the projection estimation for all bounded μ in a ball of radius L in a Besov space $B^s_{2\infty}([0,1])$. The regularity parameter s is assumed to be smaller than a known integer $(N+1)$ (see Härdle, Kerkyacharian, Picard and Tsybakov (1998)). We define

$$B(s,L) = \{f \in L_2([0,1]) \quad | \quad \max(\|f\|_\infty, \|f\|_{B^s_{2\infty}}) \le L < \infty\}.$$

Let ϕ be a scaling function of a multi-resolution analysis (a father wavelet) and ψ be the associated mother wavelet. Let, as always,

$$\phi_{jk}(x) = 2^{\frac{j}{2}}\phi(2^j x - k), \quad \psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k), k \in \mathbb{Z}, j = 0, 1, \ldots$$

As was stated above, the two functions ϕ and ψ are supposed to be compactly supported.

- *Conditions on the father wavelet function:*
  There exists a bounded non-increasing function Φ such that

$$\int_0^1 \Phi(|u|)du < \infty, \qquad |\phi(u)| \leq \Phi(|u|), \quad (a.s) \qquad \int \Phi(|u|)|u|^{N+1}du < \infty.$$

- *Conditions on the mother wavelet function:*

$$\int_0^1 \psi(u)u^k du = 0 \quad \forall k = 0, \ldots, N.$$

Let V_J be the linear subspace of $L_2([0,1])$ spanned by $\{\phi_{J,k}(x), k = 0, \ldots, 2^J - 1\}$; V_J is also spanned by the first 2^J elements of the basis $\mathcal{E} = \{\psi_{j,k}, k = 0, \ldots, 2^j - 1, j \geq -1\}$.

For each $J > 0$, we have the orthogonal projection on V_J as

$$P_{V_J}\mu(x) = P_{2^J}(\mathcal{E})\mu(x) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k}\phi_{j_0 k}(x) + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(x), \quad j_0 \geq 0.$$

Under the above conditions on the wavelet functions, we have (see Meyer (1990))

$$\|\mu - P_{2^J}(\mathcal{E})\mu\|_2 = \|\mu - P_{V_J}\mu\|_2 \leq \|\mu\|_{B^s_{2\infty}} 2^{-Js}, \quad \forall \mu \in B^s_{2,\infty}([0,1]), 0 < s < N+1.$$

Apply Theorem 3.1 with $M = 2^{J_{nm}} \asymp (nm)^{1/(2s+1)}$. We write $\hat{\mu}_{nm} = \hat{\mu}_M(\mathcal{E}) = \hat{\mu}_{2^{J_{nm}}}(\mathcal{E})$, $\delta = X - \mu$. Then, for all $\mu \in B(s, L)$, we have:

1. $E\|\hat{\mu}_{nm} - P_{V_{J_{nm}}}\mu\|_2^2 \leq C\left[\frac{2^{J_{nm}}}{nm}\left(\|\mu\|_\infty^2 + \sigma^2 + \|\omega\|_\infty\right) + \frac{2E\|\delta\|_2^2}{n}\right] \leq C_1\frac{2^{J_{nm}}}{nm} + C_2\frac{E\|\delta\|_2^2}{n}$;
2. $\|\mu - P_{V_{J_{nm}}}\mu\|_2^2 \leq L2^{-2J_{nm}s}$.

By using the wavelet basis $\mathcal{E}$, with $M = 2^{J_{nm}} \asymp (nm)^{1/(2s+1)}$, we have the following convergence properties of the wavelet estimator $\hat{\mu}_{nm}$ for all μ in a ball of radius L in a Besov space $B^s_{2\infty}([0,1])$ $(0 < s < N+1)$.

Corollary 3.1. *Under the above conditions on the wavelet functions, and the conditions of Theorem* 3.1*, for all* $\mu \in B(s, L)$*, we have*

$$\begin{aligned} E\|\hat{\mu}_{nm} - \mu\|_2^2 &= E\|\hat{\mu}_{nm} - P_{V_{J_{nm}}}\mu\|_2^2 + \|P_{V_{J_{nm}}}\mu - \mu\|_2^2 \\ &\le \left(C_1 \frac{2^{J_{nm}}}{nm} + L2^{-2J_{nm}s} \right) + C_2 \frac{E\|\delta\|_2^2}{n} \\ &\asymp \left[(nm)^{\frac{-2s}{2s+1}} + \frac{\int_0^1 \mathrm{Var}\,(X(u))du}{n} \right]. \end{aligned}$$

Comment 3.2. Corollary 3.1 gives us an upper bound on L_2 risk for the case where the recorded times T_{ij} are uniformly distributed. For the non-uniform case, where the density g of random times T_{ij} is such that $0 < g_{\min} \le g(x) \le g_{\max} < \infty$, a similar result can be obtained by assuming $E\left((\delta \circ G^{-1}(u))^2\right) \le C < \infty, \forall u \in [0, 1]$:

$$\sup_{\mu \circ G^{-1} \in B(s,L)} E\|\hat{\mu}_{nm} - \mu\|_2^2 \le \frac{C}{g_{\min}} \left((nm)^{-\frac{2s}{2s+1}} + \frac{g_{\max} \int_0^1 \mathrm{Var}(X(u))du}{n} \right),$$

where $G(x) = \int_0^x g(u)du$.

When $m >> n$, we observe n curves $X_i(u)$ of a random function $X(u) = \mu(u) + \delta(u)$. In this case, without any special conditions on X, our rates of convergence become the familiar $\int_0^1 \mathrm{Var}\,(X(u))du/n$.

When $m << n$, $m < n^{1/2s} \Leftrightarrow n^{-1} \le (nm)^{-2s/(2s+1)}$, and $C(nm)^{-2s/(2s+1)}$ plays the most important part of the rates of convergence, which corresponds to the optimal rates of the regression model. This makes sense because when $m = 1$ we have n observations $Y_i = \mu(T_i) + \epsilon_i^*$ with i.i.d. errors ϵ_i^*, so $\hat{\mu}_{nm}$ is the wavelet estimator of the regression model. Moreover, when m is a finite integer, we can prove the optimality of our wavelet estimator under Gaussian assumptions, see the next subsection.

3.3. Lower bound under Gaussian assumptions

In this subsection, we investigate the optimality of our estimator under the assumptions that m is a finite integer, the noises ϵ_{ij} are i.i.d. zero-mean Gaussian variables with finite variance $E(\epsilon_{ij}^2) < \infty$, and X is a Gaussian process.

Theorem 3.2. *Under the above assumptions,*

$$\inf_{\hat{\tau}} \sup_{\mu \in B(s,L)} E\|\hat{\tau} - \mu\|_2^2 \ge Cn^{-\frac{2s}{2s+1}}, \tag{3.3}$$

where $\inf_{\hat{\tau}}$ *indicates the infimum over the set of all possible estimators* $\hat{\tau}$ *of* μ*.*

Comment 3.3. Corollary 3.1 shows the performance of our wavelet estimator $\hat{\mu}_{nm}$. For the case where m is a finite integer and X has finite covariance, Corollary 3.1 has it that

$$\sup_{\mu\in B(s,L)} E\|\hat{\mu}_{nm}-\mu\|_2^2 \asymp (nm)^{-\frac{2s}{2s+1}} + \frac{\int_0^1 \mathrm{Var}\,(X(u))du}{n} \asymp n^{-\frac{2s}{2s+1}}.$$

Theorem 3.2 provides the lower bound for all possible estimators $\hat{\tau}$ of μ,

$$\sup_{\mu\in B(s,L)} E\|\hat{\tau}-\mu\|_2^2 \geq Cn^{-\frac{2s}{2s+1}} \geq C \sup_{\mu\in B(s,L)} E\|\hat{\mu}_{nm}-\mu\|_2^2,$$

which shows the optimality of our wavelet estimator $\hat{\mu}_{nm}$.

4. Estimation of the Covariance Function

In this section, the covariance function $\omega(u,v)$ will be estimated from the data (Y_{ij},T_{ij}) of the model (2.1). There are two cases corresponding to known and unknown mean functions μ. We start with the known μ case. Subsection 4.1 presents the covariance projection estimator. For the unknown-mean-function case, the cross-estimator is introduced and studied in Subsection 4.2. Theorem 4.3 and Theorem 4.4 show the asymptotic performances of covariance estimators. We will see that theses performances depend on the orthogonal bases which are used in the projection procedures. Subsection 4.3 investigates the L_2 risk of these estimators when wavelet-tensor bases are used.

4.1. Projection estimation for the known-mean-function case

We want to estimate the covariance function of X from the data (Y_{ij},T_{ij}) of the model (2.1):

$$\begin{aligned} Y_{ij} &= X_i(T_{ij}) + \epsilon_{ij} \\ &= \mu(T_{ij}) + \delta_i(T_{ij}) + \epsilon_{ij}, \quad i=1,\ldots,n; j=1,\ldots,m. \end{aligned}$$

The fact μ is known implies that $Z_{ij} = Y_{ij} - \mu(T_{ij})$ are observable with $Z_{ij} = \delta_i(T_{ij}) + \epsilon_{ij}, i=1,\ldots,n; j=1,\ldots,m$. For each $i=1,\ldots,n$, we write $R_{iq} = T_{i,2q-1}, U_{iq} = Z_{i,(2q-1)}, \epsilon_{iq1} = \epsilon_{i,(2q-1)}$, $S_{iq} = T_{i,2q}, V_{iq} = Z_{i,2q}$, and $\epsilon_{iq2} = \epsilon_{i,2q}$. Then we have $l = [m/2]$ sub models

$$\begin{aligned} U_{iq} &= \delta_i(R_{iq}) + \epsilon_{iq1}, \\ V_{iq} &= \delta_i(S_{iq}) + \epsilon_{iq2}, \end{aligned}$$

with $i=1,\ldots,n$, $q=1,\ldots,l$.

Let $\Pi = \{\pi_k, k \in \mathbb{N}\}$ be an orthonormal basis of $L_2([0,1]^2)$. For all $\omega \in L_2([0,1]^2)$, we have the decomposition

$$\omega(u,v) = \sum_{k\in\mathbb{N}} c_k \pi_k(u,v), \quad c_k = \int_0^1 \int_0^1 \pi_k(u,v)\omega(u,v)dudv.$$

$P_{M_1}(\Pi)\omega$ is the orthogonal projection of ω on the subspace spanned of the first M_1 elements of Π,

$$P_{M_1}(\Pi)\omega = \sum_{k=1}^{M_1} c_k \pi_k.$$

In the known-mean-function case, the projection estimator $\hat{\omega}_{M_1}(\Pi)$ of the covariance function is written as

$$\hat{\omega}_{M_1}(\Pi)(u,v) = \sum_{k=1}^{M_1} \hat{c}_k \pi_k(u,v),$$

with the unbiased estimator of coefficient c_k being

$$\hat{c}_k = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{l}\sum_{q=1}^{l} U_{iq}V_{iq}\pi_k(R_{iq}, S_{iq}) = \frac{1}{n}\sum_{i=1}^{n} \hat{c}_{k,i}, \quad l = [\frac{m}{2}],$$

where $\hat{c}_{k,i} = (1/l)\sum_{q=1}^{l} U_{iq}V_{iq}\pi_k(R_{iq}, S_{iq})$.

Theorem 4.3. *Let $\omega(u,v) = E[\delta(u)\delta(v)]$ and $h(s,t,u,v) = E[\delta(s)\delta(t)\delta(u)\delta(v)]$ be such that*

$$|\omega(u,v)| \le \|\omega\|_\infty < \infty, \forall (u,v) \in [0,1]^2, |h(s,t,u,v)| \le \|h\|_\infty < \infty, \forall (s,t,u,v) \in [0,1]^4.$$

For any finite integer $M_1 > 0$, we have

$$\begin{aligned} &E\|\hat{\omega}_{M_1}(\Pi) - P_{M_1}(\Pi)\omega\|_2^2 \\ &\quad \le C\left(\frac{2M_1}{nm}C(\sigma,\omega,h) + \frac{\|h\|_\infty}{n}\sum_{k=1}^{M_1}\left(\int_{[0,1]^2} |\pi_k(u,v)|dudv\right)^2\right), \end{aligned} \tag{4.1}$$

where $C(\sigma,\omega,h) = 2\sigma^2\|\omega\|_\infty + \sigma^4 + \|h\|_\infty$.

Comment 4.4. In the mean estimation of Section 3, the upper bound of $E\|\hat{\mu}_{nm} - P_M(\mathcal{E})\mu\|_2^2$ does not depend on the orthonormal basis $\mathcal{E}$, the wavelet basis is used only to bound the term $E\|\mu - P_M(\mathcal{E})\mu\|_2^2$ (see Comment 3.1).

For covariance estimation, the upper bound of $E\|\hat{\omega}_{M_1}(\Pi) - P_{M_1}(\Pi)\omega\|_2^2$ depends on the basis Π, more precisely, it depends on $\sum_{k=1}^{M_1}\left(\int_{[0,1]^2} |\pi_k(u,v)|dudv\right)^2$.

In Subsection 4.3, we prove that this term can be bounded when a wavelet-tensor basis is used in the projection procedure. The choice of M_1 is also discussed in Subsection 4.3.

The mean function μ is hidden from the data (Y_{ij}, T_{ij}) of model (2.1). The best thing we can do is to replace the unknown function μ by a "good" estimator $\hat{\mu}$ and apply the projection procedure with $\hat{\mu}$. In the next subsection, in order to have a "good" estimator of μ, we introduce the cross-estimator $\hat{\mu}_{(i),M}$, and study the performance of the associated covariance estimator.

4.2. Projection estimation for the unknown-mean-function case

In Section 3, under conditions on μ and δ, we have a projection estimator $\hat{\mu}_M(\mathcal{E})(u)$ for which

$$E\|\hat{\mu}_M(\mathcal{E}) - \mu\|_2^2 \le C\left(\frac{M}{nm} + \frac{E\|\delta\|_2^2}{n}\right) + \|P_M(\mathcal{E})\mu - \mu\|_2^2 = v_{nm}^2(M).$$

Cross estimators

For each $1 \le i_0 \le n$, the projection estimation procedure of Section 3 can be applied to the data $(Y_{ij}, T_{ij}) i = 1, \ldots, n, i \ne i_0, j = 1, \ldots, m$, to get the estimator

$$\hat{\mu}_{(i_0),M}(u) = \sum_{k=1}^{M} \hat{\alpha}_{(i_0),k} e_k, \quad \hat{\alpha}_{(i_0),k} = \frac{1}{(n-1)m} \sum_{i=1, i \ne i_0}^{n} \sum_{j=1}^{m} Y_{ij} e_k(T_{ij}).$$

The asymptotic behaviour of $\hat{\mu}_{(i_0),M}$ is the same as that of $\hat{\mu}$:

$$E\|\hat{\mu}_{(i_0),M} - \mu\|_2^2 \le v_{(n-1)m}^2(M) \asymp v_{nm}^2(M).$$

Return to the model (2.1) where $Z_{ij} = Y_{ij} - \mu(T_{ij}) = \delta_i(T_{ij}) + \epsilon_{ij}$ with the Z_{ij} not observable. For each i, by replacing $\{\mu(T_{ij})\}_{ij}$ by known values $\{\hat{\mu}_{(i),M}(T_{ij})\}_{ij}$, $j = 1, \ldots, m$, we have approximative values Z_{ij}^* of Z_{ij},

$$\begin{aligned} Z_{ij}^* &= Y_{ij} - \hat{\mu}_{(i),M}(T_{ij}) \approx Z_{ij} = \delta_i(T_{ij}) + \epsilon_{ij}; \\ &= Z_{ij} + \mu(T_{ij}) - \hat{\mu}_{(i),M}(T_{ij}) = Z_{ij} + \Delta_i(T_{ij}), \end{aligned}$$

where $\Delta_i(u) = \mu(u) - \hat{\mu}_{(i),M}(u)$.

In the same way, we have the approximative values (U_{iq}^*, V_{iq}^*), with

$$\begin{aligned} U_{iq}^* &= Z_{i,2q-1}^* = Z_{i,2q-1} + \Delta_i(R_{iq}) = U_{iq} + \Delta_i(R_{iq}), \quad U_{iq}^* \approx \delta_i(R_{iq}) + \epsilon_{iq1}, \\ V_{iq}^* &= Z_{i,2q}^* = Z_{i,2q} + \Delta_i(S_{i1}) = V_{iq} + \Delta_i(S_{iq}), \quad V_{iq}^* \approx \delta_i(S_{iq}) + \epsilon_{iq2}, \end{aligned}$$

for all $q = 1, \ldots, l$ with $l = [m/2]$. We remark that U_{iq}^*, V_{iq}^* are observable and U_{iq}, V_{iq} are not.

In this subsection, we consider (Π) as a tensor basis of $L_2([0,1]^2)$. The projection estimator $\hat{\omega}^*_{M_1}(\Pi)(u,v)$ is written as

$$\hat{\omega}^*_{M_1}(\Pi)(u,v) = \sum_{k=1}^{M_1} \hat{c}^*_k \pi_k(u,v), \quad \pi_k(u,v) = \pi_{k1}(u)\pi_{k2}(v),$$

where

$$\hat{c}^*_k = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{l}\sum_{q=1}^{l} U^*_{iq} V^*_{iq} \pi_k(R_{iq}, S_{iq}), \qquad l = \left[\frac{m}{2}\right].$$

Theorem 4.4. *Let $\mathcal{E} = \{e_i\}_{i\in\mathbf{N}}$ be an orthonormal basis of $L_2([0,1])$. Under the conditions of Theorem 4.3, we have*

$$\begin{aligned} E\|\hat{\omega}^*_{M_1}(\Pi) - \hat{\omega}_{M_1}(\Pi)\|_2^2 \le & \sum_{k=1}^{M_1} \left(W(\pi_{k2}, \pi_{k1}) + W(\pi_{k2}, \pi_{k1})\right) v^2_{nm}(M) \\ & + \sum_{k=1}^{M_1} C\Big(\frac{4\|\pi_k^2\|_\infty}{nm} + 1\Big) v^*_n, \end{aligned} \tag{4.2}$$

where

$$W(f,g) = \frac{2\|f^2\|_\infty}{nm} + \Big(\int_0^1 |f(u)|du\Big)\|f\|_\infty\Big(\int_0^1 |g(v)|dv\Big)^2,$$

$$v^2_{nm}(M) = C\left(\frac{M}{nm} + \frac{E\|\delta\|_2^2}{n}\right) + \|P_M(\mathcal{E})\mu - \mu\|_2^2,$$

$$v^*_n = M\sum_{t=1}^{M}\left(\frac{\int_0^1 e_t^4(u)du}{n^3} + \frac{1}{n^2}\right) + \|P_M(\mathcal{E})\mu - \mu\|_2^4.$$

Comment 4.5. Theorem 4.4 gives us an upper bound of $E\|\hat{\omega}^*_{M_1}(\Pi) - \hat{\omega}_{M_1}(\Pi)\|_2^2$ dependent on the tensor basis Π (as in Theorem 4.3). Replacing the unknown mean function μ by its cross-estimator $\hat{\mu}_{(i),M}$ implies that the upper bound depends on the performance of $\hat{\mu}_{(i),M}$ and also the basis $\mathcal{E}$ (more precisely, it contains $v^2_{nm}(M)$ and v^*_n).

The choice of M and M_1 depend on the bases $\mathcal{E}$ and Π. This choice is discussed in the next subsection where we see that the complicated terms of the upper bound lead to simple rates of convergence.

4.3. Wavelet estimation for the covariance function

In this subsection we illustrate the results of Theorem 4.3 and Theorem 4.4 with the case where the basis (Π) is a compactly supported wavelet-tensor basis on $[0,1]^2$.

To simplify the notation we suppose that the wavelet basis is described in the following way: $\Pi = \{\psi^*_{j,k}, j \in \mathbb{N}, k \in A_j\}$ where, for each $j \in \mathbb{N}$, A_j is a set with cardinality of order 2^{2j}. We have

$$\int_{[0,1]^2} |\psi^*_{j,k}(u,v)| dudv \leq C_1 2^{-j}, \|(\psi^*_{j,k})^2\|_\infty = C_2 2^{2j} \text{ with } C_1, C_2 < \infty$$

We recall some facts that hold under classical properties of regularity and vanishing moments (see Meyer (1990)).

For all functions

$$\omega = \sum_{j=0}^{\infty} \sum_{k \in A_j} \beta^*_{j,k} \psi^*_{j,k}, \omega \in L_2([0,1]^2),$$

$$\|\omega - P_{V_{J_1}}\omega\|_2 = \|\omega - \sum_{j \leq J_1} \sum_{k \in A_j} \beta^*_{j,k}\psi^*_{j,k}\|_2 \leq \|\omega\|_{B^\gamma_{2,\infty}([0,1]^2)} 2^{-J_1\gamma}. \quad (4.3)$$

Apply Theorem 4.3, with $M_1 \asymp 2^{2J_1} \asymp (nm)^{1/(1+\gamma)}$ such that $P_{M_1}(\Pi) = P_{V_{J_1}}$, and by writing $\hat{\omega}_{nm} = \hat{\omega}_{M_1}(\Pi)$, we have:

$$E\|\hat{\omega}_{nm} - P_{V_{J_1}}\omega\|_2^2 \leq C\left(\frac{2^{2J_1}}{nm} + \frac{1}{n}\sum_{j \leq J_1}\sum_{k \in A_j} 2^{-2j}\right) = C(\frac{2^{2J_1}}{nm} + \frac{J_1}{n}). \quad (4.4)$$

From (4.3) and (4.4), we have the following result.

Corollary 4.2(Known-mean-function case). *Under the conditions of Theorem* 4.3, *and assuming classical properties of regularity and moment vanishing of wavelet functions* $\psi^*_{j,k}$, *for all*

$$\omega \in B_2(\gamma, L) = \{\omega \in L_2([0,1]^2) \quad | \quad \max(\|\omega\|_\infty, \|\omega\|_{B^\gamma_{2,\infty}([0,1]^2)}) \leq L < \infty\}$$

when $m^\gamma << n$, *by taking* $2^{2J_1} \asymp (nm)^{1/(\gamma+1)}$ *we have*

$$\begin{aligned} E\|\hat{\omega}_{nm} - \omega\|_2^2 &\leq 2(E\|\hat{\omega}_{nm} - P_{V_{J_1}}\omega\|_2^2 + \|P_{V_{J_1}}\omega - \omega\|_2^2) \\ &\leq C\left(\frac{2^{2J_1}}{nm} + 2^{-2J_1\gamma}\right) \asymp (nm)^{-\frac{\gamma}{\gamma+1}}. \end{aligned}$$

Comment 4.6. By using a wavelet-tensor basis, the L_2 risk of our estimator can be bounded by $C(nm)^{-\gamma/(\gamma+1)}$, which depends on the number of experimental units (n), on the number of observations per unit (m), and on the regularity parameter of the covariance function (γ).

In practice, when longitudinal data are studied, the number of observations per unit (m) is usually a small integer, so $m^{\gamma} << n$.

The covariance function ω is often assumed to be two-times differentiable for the convergence of estimators. In our estimation procedure, the regularity parameter γ can take any positive value and we have the convergence of $\hat{\omega}_{nm}$.

We write $W(\pi_k) = W(\pi_{k1}, \pi_{k2}) + W(\pi_{k2}, \pi_{k1})$, so

$$W(\psi^*_{j,k}) \le C(\frac{2^j}{nm} + 2^{-j}) = C2^{-j}(\frac{2^{2j}}{nm} + 1).$$

For Theorem 4.4, we consider the wavelet basis $\mathcal{E}$ of $L_2([0,1])$ as in Corollary 3.1, we have

$$\int_0^1 e_t^4(u)du = \int_0^1 \psi_{jk}^4(u)du \le C2^j, \quad C < \infty.$$

With $M = 2^J$ and under classical properties of regularity and moment vanishing of wavelet functions, we have

$$v^2_{nm}(M) = C(\frac{2^J}{nm} + \frac{C_1}{n}) + \|P_{V_J}\mu - \mu\|_2^2 \le C\left(\frac{2^J}{nm} + \frac{C_1}{n} + 2^{-2Js}\right)$$

$$v^*_n = C\left(2^J \sum_{j\le J}\sum_{k=0}^{2^j-1}\left[\frac{2^j}{n^3} + \frac{1}{n^2}\right]\right) + \|P_{V_J}\mu - \mu\|_2^4 \le C\left(\frac{2^{3J}}{n^3} + \frac{2^{2J}}{n^2} + 2^{-4Js}\right).$$

When m is a finite integer, by taking $2^J \asymp n^{1/(2s+1)}$, we have $v^2_{nm}(M) \le Cn^{-2s/(2s+1)}$, $v^*_n \le Cn^{-4s/(2s+1)}$. Applying Theorem 4.4, by taking $2^{2J_1} < Cn$, we have

$$\begin{aligned} E\|\hat{\omega}^*_{M_1}(\Pi) - \hat{\omega}_{M_1}(\Pi)\|_2^2 &\le C\sum_{j\le J_1}\sum_{k\in A_j}\left[2^{-j}\left(\frac{2^{2j}}{nm}+1\right)n^{-\frac{2s}{2s+1}} + \left(\frac{2^{2j}}{nm}+1\right)n^{-\frac{4s}{2s+1}}\right] \\ &\le C\left(2^{J_1}n^{-\frac{2s}{2s+1}} + 2^{2J_1}n^{-\frac{4s}{2s+1}}\right) \\ &\le C\left(2^{J_1}n^{-\frac{2s}{2s+1}} + \frac{2^{2J_1}}{n}\right), \ \textit{with } s \ge \frac{1}{2}. \end{aligned}$$

Corollary 4.3(Unknown-mean-function case). *Let m be a finite integer. Under the conditions of Theorem 4.4, and classical properties of regularity and moment vanishing of wavelet functions, $\forall \mu \in B(s, L_1), s \ge 1/2, \forall \omega \in B_2(\gamma, L_2)$, we have*

$$\begin{aligned} E\|\hat{\omega}^*_n - \omega\|_2^2 &\le 3(E\|\hat{\omega}^*_n - \hat{\omega}_{M_1}(\Pi)\|_2^2 + E\|\hat{\omega}_{M_1}(\Pi) - P_{V_{J_1}}\omega\|_2^2 + \|P_{V_{J_1}}\omega - \omega\|_2^2) \\ &\le C\left(\frac{2^{J_1}}{n^{\frac{2s}{2s+1}}} + \frac{2^{2J_1}}{n} + 2^{-2J_1\gamma}\right) \asymp Cn^{-\tau}, \end{aligned}$$

with $\tau = \gamma/(\gamma+1)$*, if* $s > 1/2+\gamma$*, by taking* $M_1 \asymp 2^{2J_1} \asymp n^{1/(\gamma+1)}$*;* $\tau = [2s/(2s+1)][2\gamma/(2\gamma+1)]$*, if* $1/2 \leq s < 1/2+\gamma$*, by taking* $M_1 \asymp 2^{2J_1} \asymp n^{[2s/(2s+1)][2/(2\gamma+1)]}$*.*

Comment 4.7. If μ is much smoother than ω, $s > 1/2+\gamma$, the fact of replacing the unknown mean function by its wavelet estimator does not affect the performance of the covariance estimation procedure ($n^{-\gamma/(\gamma+1)}$).

When m is a finite integer, in order to estimate the mean function μ, view the model (2.1) as a regression model. Then the assumption $s \geq 1/2$ is quite natural.

5. Proof

We are interested in the quadratic risk of linear estimators. The orthogonal projection method controls quadratic risk with estimated coefficients. The main idea of our proofs is that each estimated coefficient can be decomposed into several terms that are processed differently. We present a very short version of the proof of Theorem 3.1 to illustrate this idea. One can see that the main difficulties usually come from the "δ" term-the term containing the random process $\delta = X - E(X)$. The "functional principal components expansion" of X (or δ), presented in Section 2.1, is useful in the analysis.

The complete proofs of our main theorems can be found in the on-line version of this paper at: `http://www.stat.sinica.edu.tw/statistica`.

Mean estimation - upper bound

Write

$$\begin{aligned}
A_{ki1} &= \frac{1}{m}\sum_{j=1}^{m}\mu(T_{ij})e_k(T_{ij}) - \alpha_k,\\
A_{ki2} &= \frac{1}{m}\sum_{j=1}^{m}\delta_i(T_{ij})e_k(T_{ij}) = \sum_{l\geq 1}\xi_{il}\frac{|\sum_{j=1}^{m}e_k(T_{ij})\eta_l(T_{ij})|}{m},\\
A_{ki3} &= \frac{1}{m}\sum_{j=1}^{m}\epsilon_{ij}e_k(T_{ij}).
\end{aligned}$$

So that $E(A_{ki1}^2) \leq \|\mu\|_\infty^2/m, \quad E(A_{ki3}^2) \leq \sigma^2/m$, and also

$$\begin{aligned}
\frac{E(A_{ki2}^2)}{2} \leq\, & E\Big[\Big(\sum_{l\geq 1}\xi_{il}\sum_{j=1}^{m}\frac{e_k(T_{ij})\eta_l(T_{ij}) - < e_k, \eta_l >}{m}\Big)^2\Big]\\
&+E\Big[\Big(\sum_{l\geq 1}\xi_{il} < \eta_l, e_k >\Big)^2\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{l\ge 1} \lambda_l^2 E\left[()^2\right] + \sum_{l\ge 1} \lambda_l^2 < \eta_l, e_k >^2 \\
&= \sum_{l\ge 1} \left[\lambda_l^2 \frac{\sum_{j=1}^m E\left(e_k^2(T_{ij})\eta_l^2(T_{ij})\right)}{m^2} + \lambda_l^2 < \eta_l, e_k >^2 \right] \\
&= \frac{1}{m}\int_0^1 e_k^2(t)\left[\sum_{l\ge 1}\lambda_l^2\eta_l^2(t)\right]dt + \sum_{l\ge 1}\lambda_l^2 < \eta_l, e_k >^2 \\
&\le \frac{\|\omega\|_\infty}{m} + \sum_{l\ge 1}\lambda_l^2 < \eta_l, e_k >^2 .
\end{aligned}
$$

Using Parserval's equality, we have

$$
\begin{aligned}
&E\|\hat{\mu}_M(\mathcal{E}) - P_M(\mathcal{E})\mu\|_2^2 \\
&\quad = \frac{1}{n^2}\sum_{i=1}^n E\|\hat{\mu}_{Mi}(\mathcal{E}) - P_M(\mathcal{E})\mu\|_2^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^M E\left((\alpha_{ki}-\alpha_k)^2\right) \\
&\quad = \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^M E(A_{ki1}^2) + E(A_{ki2}^2) + E(A_{ki3}^2) \\
&\quad \le \frac{M}{nm}(\|\mu\|_\infty^2 + \sigma^2 + 2\|\omega\|_\infty) + 2\frac{1}{n}\sum_{k=1}^M\sum_{l\ge 1}\lambda_l^2 < \eta_l, e_k >^2 \\
&\quad = \frac{M}{nm}(\|\mu\|_\infty^2 + \sigma^2 + 2\|\omega\|_\infty) + 2\frac{E\|P_M(\mathcal{E})\delta\|_2^2}{n}.
\end{aligned}
$$

Acknowledgement

I would like to thank Dominique Picard for her thorough and useful comments which have helped to improve the presentation of the paper.

References

Boularan, J., Ferré, L., and Vieu, P. (1995). A nonparametric model for unbalanced longitudinal data, with application to geophysical data. *Comput. Statist.* **10**, 285-298.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Applications.* Chapman and Hall, London.

Grenander, U. (1968). *Stochastic Processes and Statistical Inference.* 2nd edition. Almqvist and Wiksell, Stockholm.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). Wavelet, Approximation and Statistical Applications. Lecture Notes in Statistics. 129, Springer Verlag, New York.

Indritz, J. (1963). *Methods in Analysis.* Macmillan, New York.

Meyer, Y. (1990). *Ondelettes et Opérateurs.* Hermann, Paris.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis.* Springer, NewYork.

Rice, J. and Silverman, B. W. (1991). Estimating the mean and covariance structure non parametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53**, 233-243.

Tsybakov, A. B. (2004). *Introduction a l'Estimation Non-parametrique.* Springer.

Yao, F., Müller, H. G., and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.

Laboratoire de Probabilités et Modèles Aléatoires, LPMA, Université Paris VII (Denis Diderot), 175 rue du Chevaleret, 75013 Paris, France.

E-mail: nguyen@math.jussieu.fr

(Received June 2007; accepted December 2007)