Ridge regression for the functional concurrent model

Tito Manrique Chuquillanqui, Christophe Crambes, Nadine Hilgert

Electronic Journal of Statistics
Vol. 12 (2018) 985–1018
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1412

Ridge regression for the functional concurrent model

Tito Manrique*

MISTEA, INRA, Montpellier SupAgro, Univ. Montpellier, Montpellier, France
and IMAG, Univ. Montpellier, Montpellier, France
e-mail: tkarl.manrique@protonmail.com

Christophe Crambes

IMAG, Univ. Montpellier, Montpellier, France
e-mail: christophe.crambes@umontpellier.fr

and

Nadine Hilgert

MISTEA, INRA, Montpellier SupAgro, Univ. Montpellier, Montpellier, France
e-mail: nadine.hilgert@inra.fr

Abstract: The aim of this paper is to propose estimators of the unknown functional coefficients in the Functional Concurrent Model (FCM). We extend the Ridge Regression method developed in the classical linear case to the functional data framework. Two distinct penalized estimators are obtained: one with a constant regularization parameter and the other with a functional one. We prove the probability convergence of these estimators with rate. Then we study the practical choice of both regularization parameters. Additionally, we present some simulations that show the accuracy of these estimators despite a very low signal-to-noise ratio.

MSC 2010 subject classifications: Primary 62J05, 62G05, 62G20; secondary 62J07.
Keywords and phrases: Functional linear model, functional data, ridge regression, concurrent model, varying coefficient model.

Received May 2017.

Contents

*The research of Tito Manrique was supported in part by the Labex NUMEV (convention ANR-10-LABX-20) under project 2013-1-007.

1. Introduction

Functional Data Analysis (FDA) proposes very good tools to handle data that are functions of some covariate (e.g. time, when dealing with longitudinal data), see Hsing and Eubank [11] or Horváth and Kokoszka [10]. These tools allow for better modelling of complex relationships than classical multivariate data analysis do, as noticed by Ramsay and Silverman [15, Ch. 1], Yao et al. [20, 19], among others.

There are several models in FDA for studying the relationship between two variables. In particular in this paper we are interested in the Functional Concurrent Model (FCM) which is defined as follows

$$\mathcal{Y}(t) = \beta_0(t) + \beta_1(t)\,\mathcal{X}(t) + \epsilon(t), \tag{1.1}$$

where $t \in \mathbb{R}$, β_0 and β_1 are the unknown functions to be estimated, $\mathcal{X}, \mathcal{Y}$ are random functions and ϵ is a noise random function. All the functions considered here are complex valued.

From a practical perspective all functional linear models can be reduced to a functional concurrent model with several covariates (Ramsay and Silverman [15, p. 220]). This model is also related to the functional varying coefficient model (VCM) and has been studied for example by Wu et al. [18] or more recently by Şentürk and Müller [16].

Another practical advantage of model (1.1) is that it allows to simplify the study of the following convolution model

$$W(s) = \int_{-\infty}^{+\infty} \theta(u)\, Z(s-u)du + \eta(s), \tag{1.2}$$

where $u, s \in \mathbb{R}$, through the Fourier transform $\mathcal{F}$ with $Y = \mathcal{F}(W)$, $\beta_0 \equiv 0$, $\beta_1 = \mathcal{F}(\theta)$, $X = \mathcal{F}(Z)$ and $\epsilon = \mathcal{F}(\eta)$.

Despite the abundant literature related to FCM or functional VCM, there is hardly any paper providing estimators of the unknown functions in model (1.1) along with their asymptotic properties which use the norm of the functional space where they belong to.

As noticed by Ramsay and Silverman [15, p. 259], most of the current methods of estimation come from a multivariate data analysis approach rather than from a functional one. For some applications, for example when the observations are highly auto-correlated, taking this functional nature into account may be decisive. If not, multivariate approaches may cause a loss of information because, as noticed by Şentürk and Müller [16, p. 1257], they "do not take full advantage of the functional nature of the underlying data". In practice this loss of information may reduce the accuracy of estimation and prediction. To circumvent this problem, Şentürk and Müller [16] propose a three-step functional approach based on smoothing and least square estimation. However, the convergence results obtained on compact sets do not allow to study specific models like (1.2), for which convergence on the whole real line is required.

Besides, Ramsay et al. [14, Ch 10] propose a practical estimation method by projecting all the random functions to an adequate finite dimensional subspace and then use a penalization to chose the estimator. They do not provide a theoretical study of its asymptotic properties.

The objective of the present paper is to propose estimators of the functions β_0 and β_1 in the FCM (1.1) for which the asymptotic properties are obtained. Our estimation approach is based on the Ridge Regression method developed in the classical linear case, see Hoerl [8]. We extend this to the functional data framework of model (1.1).

To ease the notations and the presentation of the results, we introduce in section 2 a simplified centered model. The functional ridge regression estimator of the functional coefficient is then defined with a constant regularization parameter. In section 3 we establish the consistency of this estimator and get a rate of convergence. Section 4 addresses the practical choice of the regularization parameter through cross-validation criteria. We also introduce a more flexible estimator with a functional regularization parameter. Some simulation trials are presented in section 5, and show the comparison of the two penalized estimators together with that of Ramsay et al. [14, Ch 10], in a very low signal-to-noise ratio (SNR) setting. Finally an application on a real data set is presented in section 6. All the proofs are postponed to Section 8.

2. Estimator and hypotheses

Let $(\mathcal{X}_i, \mathcal{Y}_i)_{i=1,\cdots,n}$ be an i.i.d sample of FCM (1.1). To remove the functional intercept β_0, we center the model (1.1) and get

$$\mathcal{Y}(t) - \mathbb{E}[\mathcal{Y}](t) \;=\; \beta_1(t)\,(\mathcal{X}(t) - \mathbb{E}[\mathcal{X}](t)) + \epsilon(t).$$

The estimator of β_0 depends on the estimator of β_1 obtained from the centered model. Given that the natural estimators of $\mathbb{E}[\mathcal{X}]$ and $\mathbb{E}[\mathcal{Y}]$ are the empirical means ($\bar{\mathcal{X}}_n := 1/n \sum_{i=1}^n \mathcal{X}_i$ and $\bar{\mathcal{Y}}_n := 1/n \sum_{i=1}^n \mathcal{Y}_i$), the estimator of β_0 is defined as

$$\hat{\beta}_0 := \bar{\mathcal{Y}}_n - \hat{\beta}_1\,\bar{\mathcal{X}}_n. \tag{2.1}$$

The convergence results on $\hat{\beta}_1$ immediately transpose to $\hat{\beta}_0$. Now, to focus on the estimation of β_1, we define the elements of the centered model as follows, $X := \mathcal{X} - \mathbb{E}[\mathcal{X}]$, $Y := \mathcal{Y} - \mathbb{E}[\mathcal{Y}]$ and $\beta := \beta_1$ and the centered FCM writes

$$Y(t) = \beta(t)\,X(t) + \epsilon(t). \tag{2.2}$$

In what follows we discuss the estimation of β.

2.1. Functional ridge regression estimator (FRRE)

The definition of the estimator of β in the centered model (2.2) is inspired by the estimator introduced by Hoerl [8] in the Ridge Regularization method that deal with ill-posed problems in the classical linear regression. Let $\lambda_n > 0$ be a regularization parameter, we define the Functional Ridge Regression Estimator (FRRE) of β as follows

$$\hat{\beta}_n := \frac{\frac{1}{n}\sum_{i=1}^n Y_i\,X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 \;+\; \frac{\lambda_n}{n}}, \tag{2.3}$$

where the exponent $*$ stands for the complex conjugate. In the classical linear regression case, Hoerl and Kennard [9, p. 62] proved that there is always a regularization parameter for which the ridge estimator is better than the Ordinary Linear Squares (OLS) estimator. Huh and Olkin [12] made a study of some asymptotic properties of the ridge estimator in this case. In the context of the functional linear regression with scalar output, Hall et al. [6, p. 73] have also used a ridge regularization method to invert the whole covariance operator of X. Their approach has two main differences with the one used to define the FRRE: we use i) functional outputs (Y_i) and ii) inversion of the diagonal terms of the covariance matrix of X.

In our case, the use of λ_n in the denominator prevents from dividing by zero because $\mathbb{E}[X] = 0$ (centered model) and, therefore, it helps to control the instability of the estimator. The simulation studies in Section 5 show that in practice a better estimator is obtained with the regularization parameter.

2.2. Notations and general hypotheses of the FCM

Before studying the FCM, let us define some useful notations. We define $L^2(\mathbb{R},\mathbb{C}) = L^2$ the set of square integrable complex valued functions, with the L^2-norm $\|f\|_{L^2} := \left[\int_{\mathbb{R}} |f(x)|^2 dx\right]^{1/2}$, with its associated inner product $\langle\cdot,\cdot\rangle$. Besides, given a subset $K \subset \mathbb{R}$, $\|f\|_{L^2(K)} := \left[\int_K |f(x)|^2 dx\right]^{1/2}$, where $|\cdot|$ denotes the complex modulus.

The theoretical results given in the next sections are proved on the whole real line. For this reason, we need to restrict the study to the set of functions that vanish at infinity. Let $C_0(\mathbb{R},\mathbb{C}) = C_0$ be the space of complex valued continuous functions, which satisfies: for all $\zeta > 0$ there exists a $R > 0$ such that for all $|t| > R$, $|f(t)| < \zeta$. We use the supremum norm $\|f\|_{C_0} := \sup_{x\in\mathbb{R}} |f(x)|$. In particular for a subset $K \subset \mathbb{R}$, $\|f\|_{C_0(K)} := \sup_{x\in K} |f(x)|$.

Finally, throughout this paper, the support of a continuous function $f : \mathbb{R} \to \mathbb{C}$ is the set $supp(f) := \{t \in \mathbb{R} : |f(t)| \neq 0\}$. This set is open because f is continuous. Besides we define the boundary of a set S, as $\partial(S) := \overline{S} \setminus int(S)$, where $\overline{S}$ is the closure of S and $int(S)$ is its interior.

The space C_0 is too large. For instance, its geometry does not allow for the application of the Central Limit Theorem (CLT) under the general hypothesis of the existence of the covariance operator, that is $\mathbb{E}(\|X\|^2_{C_0}) < \infty$ (see Ledoux and Talagrand [13, Ch 10]). To circumvent this difficulty, we consider functions that belong to the space $C_0 \cap L^2$. Here are general hypotheses that will be used all along the paper:

$(HA1_{FCM})$	X, ϵ are independent $C_0 \cap L^2$ valued random functions, such that $\mathbb{E}(\epsilon) = 0$,
$(HA2_{FCM})$	$\beta_0, \beta_1 \in C_0 \cap L^2$,
$(HA3_{FCM})$	$\mathbb{E}(\|\epsilon\|^2_{C_0}), \mathbb{E}(\|X\|^2_{C_0})$, $\mathbb{E}(\|\epsilon\|^2_{L^2})$ and $\mathbb{E}(\|X\|^2_{L^2})$ are all finite.

We do not assume that $\mathbb{E}[X] = 0$ in model (2.2). Therefore, we deal with a more general case than the one derived after centering model (1.1), and our results will be valid also for the centered case.

3. Asymptotic properties of the FRRE

From the definition (2.3), it is easy to show that the FRRE $\hat{\beta}_n$ has the following bias-variance decomposition:

$$\hat{\beta}_n = \beta - \frac{\lambda_n}{n}\left[\frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}\right] + \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}. \tag{3.1}$$

In this equation, we can see that the penalization introduces a bias but helps to control the variance (last term in (3.1)). Thus, the penalization should not be too big nor too small. Note also that, when $\mathbb{E}[X] \approx 0$, the part of the denominator $\frac{1}{n}\sum_{i=1}^n |X_i(t)|^2$ might be close to zero at some values of t. Therefore, the penalization ($\lambda_n > 0$) is necessary to prevent the denominator to be too small.

Clearly, $\frac{1}{n}\sum_{i=1}^{n}|X_i(t)|^2$ is an estimator of $\mathbb{E}[|X|^2]$. Then from the equation (3.1) we deduce that the ill-posed degree of this estimation problem depends on the intervals where these two conditions are satisfied: i) $\mathbb{E}[|X|^2]$ is close to zero and ii) $\frac{\beta}{\mathbb{E}[|X|^2]}$ is not close to zero. The latter condition implies a big bias because β will be significantly bigger than the denominator.

The main results of this paper are the probability convergence of the FRRE with rate

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right),$$

and the mean square error rate

$$\mathbb{E}(\|\hat{\beta}_n - \beta\|_{L^2}^2) = O\left(\max\left[\frac{\lambda_n^2}{n^2}, \frac{n}{\lambda_n^2}\right]\right),$$

under large conditions.

3.1. Consistency of the estimator

Theorem 3.1. *Let us consider the FCM with the general hypotheses ($HA1_{FCM}$), ($HA2_{FCM}$) and ($HA3_{FCM}$). Let $(X_i, Y_i)_{i\geq 1}$ be i.i.d. realizations. We suppose moreover that*

(A1) $\overline{supp(|\beta|)} \subseteq \overline{supp(\mathbb{E}[|X|])}$,
(A2) $(\lambda_n)_{n\geq 1} \subset \mathbb{R}^+$ *is such that* $\frac{\lambda_n}{n} \to 0$ *and* $\frac{\sqrt{n}}{\lambda_n} \to 0$ *as* $n \to +\infty$.

Then

$$\lim_{n\to+\infty} \|\hat{\beta}_n - \beta\|_{L^2} = 0 \quad \text{in probability.} \tag{3.2}$$

Remark 3.2. *The geometry of L^2 helps in the proof of Theorem 3.1 to use the Central Limit Theorem (CLT) (see Bosq [1, p. 53]). In this sense by paying attention to the geometry of L^p spaces, for some $p \geq 1$, it is also possible to generalize this result for those spaces.*

Remark 3.3. *Hypothesis (A2) is classic in the context of ridge regression. Hypothesis (A1) specifies that it is not possible to estimate β outside the support of the modulus of X. From model (2.2), it is clear that β cannot be estimated in the intervals where the function X is zero. We show this in Proposition 3.4, where Hypothesis (nA1) is stronger than the negation of (A1). This hypothesis provides that there exists some t_0 in $\overline{supp(|\beta|)}$, such that X is zero almost surely in a neighborhood of t_0.*

Proposition 3.4. *Let $(X_i, Y_i)_{i=1,\cdots,n}$ be an i.i.d. sample of FCM in $C_0 \cap L^2$ which satisfies hypothesis (A2) and*

(nA1) There exists $t_0 \in \overline{supp(|\beta|)}$ and $\delta > 0$ such that $\mathbb{E}[\|X\|_{C_0([t_0-\delta,t_0+\delta])}] = 0$.

Then there exists a constant $C > 0$ such that almost surely

$$\|\hat{\beta}_n - \beta\|_{L^2} \geq C. \tag{3.3}$$

In what follows we obtain some rates of convergence over the whole real line and over compact subsets.

3.2. Rate of convergence

To obtain a rate of convergence, we need to control the shapes of the functions β and $\mathbb{E}[|X|]$ on the points at the border of the support of $\mathbb{E}[|X|]$. Theorem 3.5 handles the general case where $|\beta|/\mathbb{E}[|X|^2]$ goes to infinity over the points of the set $C_{\beta,\partial X} := supp(|\beta|) \cap \partial(supp(\mathbb{E}[|X|]))$.

Theorem 3.5. *Let us consider the FCM with the general hypotheses ($HA1_{FCM}$), ($HA2_{FCM}$) and ($HA3_{FCM}$). We assume additionally that* (A1) *holds, together with:*

(A3) $\mathbb{E}[\||X|^2\|_{L^2}^2] < \infty$.

(A4) $\left\| \frac{|\beta|}{\mathbb{E}[|X|^2]} 1_{\overline{supp(\beta)} \setminus \partial(supp(\mathbb{E}[|X|]))} \right\|_{L^2} < +\infty$.

(A5) There exist positive real numbers $\alpha > 0$, $M_0, M_1, M_2, L_I > 0$ *such that for every* $p \in C_{\beta,\partial X}$, *there exists an open neighborhood* $J_p \subset supp(|\beta|)$ *with length* $m(J_p) < L_I$ *for which the following hold*

(a) For every $t \in J_p$, $\mathbb{E}[|X|^2(t)] \geq |t-p|^\alpha$, *and*

$$\left\| \frac{1}{\mathbb{E}[|X|^2]} \right\|_{L^2(J_p \setminus \{p\})} \leq M_0,$$

(b) $\sum_{p \in C_{\beta,\partial X}} \|\beta\|_{C_0(J_p)}^2 < M_1$,

(c) $\frac{|\beta|}{\mathbb{E}[|X|^2]} 1_{supp(|\beta|) \setminus J} < M_2$, *where* $J = \bigcup_{p \in C_{\beta,\partial X}} J_p$.

(A6) For $n \geq 1$,

$$\lambda_n := n^{1 - \frac{1}{4\alpha+2}},$$

where $\alpha > 0$ *comes from the hypothesis* (A5).

Then

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(n^{-\gamma}\right), \tag{3.4}$$

where $\gamma := \min\left[\frac{1}{2(2\alpha+1)}, \frac{1}{2} - \frac{1}{2(2\alpha+1)}\right]$ *and* $n^{-\gamma} = \max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]$.

The following corollary specifies the rate of convergence for $\alpha = 1/2$.

Corollary 3.6. *Under the hypotheses of Theorem 3.5 if* $\alpha = 1/2$ *we have*

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(n^{-1/4}\right).$$

Remark 3.7. *Hypothesis* (A3) *is classic and allows to apply the CLT on the denominator of* $\hat{\beta}_n$. *Hypothesis* (A4) *is needed because* $\left[\frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}\right]$ *in (3.1) can naturally be* L^2*-bounded under this condition.*

Next (A5a) *requires that around the points* $p \in C_{\beta,\partial X}$, *the function* $\mathbb{E}[|X|^2]$ *goes to zero slower than a polynomial of degree* α, *which implies that the term*

$\left[\frac{\beta}{\frac{1}{n}\sum_{i=1}^{n}|X_i|^2+\frac{\lambda_n}{n}}\right]$ *in (3.1) behaves like* $\frac{\beta}{\mathbb{E}[|X|^2]}$ *and determines the rate of convergence.*

The degree of ill-posedness of the problem depends on how close to zero $\mathbb{E}[|X|^2]$ *is. The hypothesis* $(A5a)$ *measures this through the polynomial degree* α*. In this way the rate of convergence, which directly depends on* α*, is related to the ill-posed nature of the problem.*

Parts (b) *and* (c) *of* $(A5)$ *help us controlling the tails of* β *and* $|X|$ *around infinity. They are useful only when* $card(C_{\beta,\partial X}) = +\infty$*. Note that the set* $C_{\beta,\partial X}$ *is always countable (see the proof of Theorem 3.5).*

Finally hypothesis $(A6)$ *replaces* $(A2)$ *in Theorem 3.1, as the rate of convergence strongly depends on the behaviour of* $\frac{\beta}{\mathbb{E}[|X|^2]}$ *around the points of* $C_{\beta,\partial X}$*, which depends on* α*. We can see that* $(A6)$ *always implies* $(A2)$*.*

Remark 3.8. *It is natural to ask whether the convergence rate obtained in Theorem 3.5 is optimal or not. Stone [17] obtained an optimal convergence rate in a multivariate nonparametric regression setting. Transposition for statistical models with functional variables is still an open problem. In our case, the convergence rate in Theorem 3.5 can be written under the form* $n^{-\alpha/(2\alpha+1)}$ *with* $\alpha < 1/2$*, which leads to a rate slower than* $n^{-1/4}$*. The condition* $\alpha < 1/2$ *prevents from getting convergence rates of the same form than those given in Stone [17]. This constraint enables to bound the quantity* $\frac{1}{\mathbb{E}(|X|^2)}$*, which is crucial to control the bias term. Indeed, the convergence rate is stated in a large setting (i) for the* L^2 *norm over the whole real line, (ii) without any assumption on the regularity of the curve* X*, and (iii) without any assumption on the distribution of* X*.*

Under stronger but more intuitive hypotheses, we can also obtain similar convergence results to that of Theorem 3.5. Corollary 3.9 is an example.

Corollary 3.9. *If additionally to hypotheses* $(A1)$*,* $(A2)$ *and* $(A3)$*, we assume*

(A4bis) $\frac{|\beta|}{\mathbb{E}[|X|^2]}\,1_{supp(|\beta|)} \in L^2 \cap L^\infty$*,*

then

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{3.5}$$

Hypothesis $(A4bis)$ is a reformulation of $(A4)$ and part (c) of $(A5)$. It is required to control the second term of (3.1) and the decreasing rate of β with respect to $\mathbb{E}[|X|^2]$ around infinity (tails control). Besides, note that $(A4bis)$ implies that $C_{\beta,\partial X} = \emptyset$.

Theorem 3.10 presents a simpler convergence result on compact subsets of the support of $\mathbb{E}[|X|]$. This theorem assumes general hypotheses and ensures convergence in a wide variety of cases.

Theorem 3.10. *Under hypotheses* $(A1)$*,* $(A2)$ *and* $(A3)$*, for every compact subset* $K \subset supp(\mathbb{E}[|X|])$*, we have*

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{3.6}$$

3.3. Further results

In the previous subsection, we presented some convergence theorems that use convergence in probability (consistency). By adapting the arguments in the proof, we can also obtain convergence of the mean square error. We proved the following theorem.

Theorem 3.11. *Under hypotheses* $(A1)$, $(A2)$, $(A3)$ *and* $(A4bis)$ *we obtain*

$$\mathbb{E}[\|\hat{\beta}_n - \beta\|_{L^2}^2] = \int_{\mathbb{R}} \mathbb{E}[|\hat{\beta}_n - \beta|^2] = O\left(\max\left[\frac{\lambda_n^2}{n^2}, \frac{n}{\lambda_n^2}\right]\right). \tag{3.7}$$

Moreover, the confidence bands of β are computed in Proposition 3.12 under suitable noise conditions. We first compute the expectation and the variance of $\hat{\beta}_n$ conditionally to the sample $X_1, \cdots, X_n$. Then we define an unbiased estimator of the variance of the noise for each value $t \in \mathbb{R}$, with which we compute the confidence interval of β for this value t.

Proposition 3.12. *The expectation and variance of* $\hat{\beta}_n$ *conditional to a sample* $X_1, \cdots, X_n$ *are*

$$\mathbb{E}[\hat{\beta}_n \,|\, X_1, \cdots, X_n] = \beta D_X \quad \textit{and} \quad \mathrm{Var}[\hat{\beta}_n \,|\, X_1, \cdots, X_n] = \frac{\mathbb{E}[|\epsilon|^2]}{\sum_{i=1}^n |X_i|^2} D_X^2,$$

where D_X *is a function defined for all* $t \in \mathbb{R}$ *as follows* $D_X(t) := \frac{\frac{1}{n}\sum_{i=1}^n |X_i(t)|^2}{\frac{1}{n}\sum_{i=1}^n |X_i(t)|^2 + \frac{\lambda_n}{n}}$.

Additionally, if for a given value $t \in \mathbb{R}$, *we suppose that* $\beta(t) \in \mathbb{R}$, $X(t) \in \mathbb{R}$, $\epsilon(t) \sim N(0, \sigma_\epsilon^2)$ *and* $(\epsilon_i(t))_{i=1,\cdots,n}$ *is a i.i.d sample, then*

$$\frac{\hat{\beta}_n(t) - \beta(t) D_X(t)}{\hat{\sigma}_\epsilon D_X (\sum_{i=1}^n |X_i|^2)^{-1/2}} \sim \mathcal{T}(n-1),$$

where

$$\hat{\sigma}_\epsilon := \frac{1}{(n-1) D_X^2(t)} \sum_{i=1}^n |D_X(t) Y_i(t) - \hat{\beta}_n(t) X_i(t)|^2$$

is a unbiased estimator of $\sigma_\epsilon(t)$ *and* $\mathcal{T}(n-1)$ *is the Student's t-distribution with* $n-1$ *degrees of freedom.*

Consequently a confidence interval of $\beta(t)$ *at the level* $1-\alpha$ *is the following*

$$\begin{aligned}
&[\frac{\hat{\beta}_n(t)}{D_X(t)} - t_{n-1}(1-\alpha/2) \frac{\hat{\sigma}_\epsilon}{\sqrt{\sum_{i=1}^n |X_i(t)|^2}}, \\
&\frac{\hat{\beta}_n(t)}{D_X(t)} + t_{n-1}(1-\alpha/2) \frac{\hat{\sigma}_\epsilon}{\sqrt{\sum_{i=1}^n |X_i(t)|^2}}],
\end{aligned}$$

with critical value $t_{n-1}(1-\alpha/2)$.

4. Selection of the regularization parameter

4.1. Predictive and generalized cross-validation

This section is devoted to developing a selection procedure of the regularization parameter λ_n for a given sample $(X_i, Y_i)_{i\in\{1,\cdots,n\}}$. To solve this problem we chose the Predictive Cross-Validation (PCV) criterion. Its definition, see for instance Febrero-Bande and Oviedo de la Fuente [4, p. 17] or Hall and Hosseini-Nasab [7, p. 117], is the following

$$PCV(\lambda_n) := \frac{1}{n}\sum_{i=1}^{n} \|Y_i - \hat{\beta}_n^{(-i)}\, X_i\|_{L^2}^2,$$

where $\hat{\beta}_n^{(-i)}$ is computed with the sample $(X_j, Y_j)_{j\in\{1,\cdots,i-1,i+1,\cdots,n\}}$. The selection method consists in choosing the value λ_n which minimizes the function $PCV(\cdot)$.

Proposition 4.3 shows how to compute faster the PCV by only processing one regression, instead of n. This result is based on similar ideas as those in Green and Silverman [5, pp. 31-33] about the smoothing parameter selection for smoothing splines.

Proposition 4.1. *We have*

$$PCV(\lambda_n) = \frac{1}{n}\sum_{i=1}^{n} \left\| \frac{Y_i - \hat{\beta}_n\, X_i}{1 - A_{i,i}} \right\|_{L^2}^2, \tag{4.1}$$

where $A_{i,i} \in L^2$ is defined as follows $A_{i,i} := |X_i|^2/(\sum_{j=1}^n |X_j|^2 + \lambda_n)$.

Next we introduce the following Generalized Cross-Validation (GCV), which is computationally faster than the PCV:

$$GCV(\lambda_n) := \frac{1}{n}\sum_{i=1}^{n} \left\| \frac{Y_i - \hat{\beta}_n\, X_i}{1 - A} \right\|_{L^2}^2,$$

where $A \in L^2$ is $A := (\frac{1}{n}\sum_{i=1}^n |X_i|^2)/(\sum_{j=1}^n |X_j|^2 + \lambda_n)$.

Remark 4.2. *From the definition of A, we have that, for every $t \in \mathbb{R}$, $0 \le A(t) \le 1/n$, then $1 \le \frac{1}{1-A(t)} \le \frac{n}{n-1}$, which yields that the GCV criterion is bounded as follows:*

$$\frac{1}{n}\sum_{i=1}^{n} \left\| Y_i - \hat{\beta}_n\, X_i \right\|_{L^2}^2 \le GCV(\lambda_n) \le \frac{1}{n-1}\sum_{i=1}^{n} \left\| Y_i - \hat{\beta}_n\, X_i \right\|_{L^2}^2.$$

This last inequality quickly gives an idea about the GCV values.

4.2. Functional regularization parameter

Given that we are working with functional data, another possibility for the estimator defined in (2.3) is to use a time-dependent function $\Lambda_n(t)$ instead of

a constant number λ_n. We shall optimize, for each time t, the choice of $\Lambda_n(t)$. To that aim, we have to compute the PCV for each time $t \in \mathbb{R}$,

$$PCV(\Lambda_n(t)) := \frac{1}{n}\sum_{i=1}^{n} |Y_i(t) - \hat{\beta}_n^{(-i)}(t)\, X_i(t)|^2,$$

where $\hat{\beta}_n^{(-i)}(t)$ is computed with the sample $(X_j(t), Y_j(t))_{j\in\{1,\cdots,n\}\setminus\{i\}}$.

As above, we obtain a simpler formula for $PCV(\Lambda_n(t))$ (see next proposition bellow), which yields a faster computation.

Proposition 4.3. *We have*

$$PCV(\Lambda_n(t)) = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{Y_i(t) - \hat{\beta}_n(t)\, X_i(t)}{1 - A_{i,i}(t)} \right|^2, \tag{4.2}$$

where $A_{i,i}(t) := \frac{|X_i(t)|^2}{\sum_{j=1}^{n} |X_j(t)|^2 + \lambda_n(t)}$.

This criterion is discussed in the next section dedicated to simulation studies. Its performance is evaluated and compared to that of $GCV(\lambda_n)$.

Theoretical results can be obtained on the asymptotic properties of the estimator associated to the functional regularization parameter. For instance we proved the following theorem.

Theorem 4.4. *If additionally to the hypotheses* (*A*1), (*A*3) *and* (*A*4*bis*) *we assume*

(A2bis) There exists a constant $b > 0$ *and a set of continuous functions* $\Lambda_n : \mathbb{R} \to \mathbb{R}^+$ *such that for each* $n \in \mathbb{N}$, $M_{\Lambda_n} < b\, m_{\Lambda_n}$ *and*

$$\frac{m_{\Lambda_n}}{n} \to 0 \qquad \textit{and} \qquad \frac{\sqrt{n}}{m_{\Lambda_n}} \to 0,$$

where $m_{\Lambda_n} := \min(\Lambda_n)$ *and* $M_{\Lambda_n} := \max(\Lambda_n)$.

Then

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{m_{\Lambda_n}}{n}, \frac{\sqrt{n}}{m_{\Lambda_n}}\right]\right), \tag{4.3}$$

where $\hat{\beta}_n$ *is obtained with* $\Lambda_n(t)$ *minimizing (4.2).*

5. Simulation study

We divide the simulation study into two parts. Firstly, we present in settings 1 and 2, a comparative numerical analysis of different estimators used for estimation in model (1.1). Then, in the second part, a third setting simulation is introduced to numerical study the dependence of the convergence rate ($n^{-\gamma}$) on α, where α is a bound for the decreasing rate of $\mathbb{E}[|X|^2]$ towards 0, as described in Theorem 3.5. In this case we use the model without intercept (2.2).

5.1. Comparison of estimation methods

For settings 1 and 2, we evaluate our estimation procedures when the Signal-to-Noise-Ratio (SNR) is low, that is, under noisy conditions. Both approaches for computing the FRRE (using λ_n and $\Lambda_n(t)$) are compared along with the non penalized case ($\lambda_n = 0$). Furthermore, we also compare them to the estimator defined by Ramsay et al. ([14, Ch 10]). In this approach, the random functions are projected onto an adequate finite-dimensional subspace generated by the Fourier basis. The estimator is obtained as a solution of a penalized least square criterion and is implemented in the **R** package **fda**.

We use the estimator (2.1) of β_0 and the FRRE estimator of β_1 after centering, that is

$$\begin{aligned} \hat{\beta}_1 &:= \frac{\sum_{i=1}^{n}(\mathcal{Y}_i - \bar{\mathcal{Y}}_n)\,(\mathcal{X}_i - \bar{\mathcal{X}}_n)^*}{\sum_{i=1}^{n}|\mathcal{X}_i - \bar{\mathcal{X}}_n|^2 \;+\; \lambda_n}, \\ \hat{\beta}_0 &:= \bar{\mathcal{Y}}_n - \hat{\beta}_1\,\bar{\mathcal{X}}_n. \end{aligned} \tag{5.1}$$

For each setting we computed 500 Monte Carlo runs to evaluate the mean absolute deviation error (MADE) and the weighted average squared error (WASE), defined in the same way as in Şentürk and Müller [16, p. 1261],

$$\begin{aligned} MADE &:= \frac{1}{2T}\left[\frac{\int_0^T |\beta_0(t) - \hat{\beta}_0(t)|\,dt}{range(\beta_0)} + \frac{\int_0^T |\beta_1(t) - \hat{\beta}_1(t)|\,dt}{range(\beta_1)}\right], \\ WASE &:= \frac{1}{2T}\left[\frac{\int_0^T |\beta_0(t) - \hat{\beta}_0(t)|^2 dt}{range^2(\beta_0)} + \frac{\int_0^T |\beta_1(t) - \hat{\beta}_1(t)|^2 dt}{range^2(\beta_1)}\right], \end{aligned}$$

where $[0, T]$ is the domain of β_0 and β_1 and $range(\beta_r)$ is the range of the function β_r for $r = 0, 1$.

In the first setting, we analyze how the estimators behave when $\mathbb{E}[\mathcal{X}] > 0$. Then, in the second one, we study a case where the penalization ($\lambda > 0$) is clearly needed, that is, when $\mathbb{E}[\mathcal{X}] = 0$ and $\beta_0 = 0$.

For both settings, we simulated random functions $(\mathcal{X}_i, \mathcal{Y}_i)_{i=1,\cdots,n}$ over the interval $[0, 1]$, discretized in $p = 100$ equispaced observation times $t_j := j/101$ for $j = 1, \cdots, 100$. Additionally, to measure the level of noise, we use the signal-to-noise ratio (SNR), defined by $SNR := (tr(Cov(\mathcal{X})))/(tr(Cov(\epsilon))$, where $Cov(\mathcal{X}) := \mathbb{E}\,(< \mathcal{X}, \cdot > \mathcal{X} - < \mathbb{E}(\mathcal{X}), \cdot > \mathbb{E}(\mathcal{X}))$, $Cov(\epsilon) := \mathbb{E}\,(< \epsilon, \cdot > \epsilon)$ and tr is the trace of an operator.

The general hypotheses $(HA1_{FCM})$ - $(HA3_{FCM})$ are satisfied for both settings. The regularization parameter λ_n and the function Λ_n were optimized over the interval $[0, 100]$.

5.1.1. Setting 1

We simulated samples with size $n = 70$. The input curves $\mathcal{X}_i$, for $i = 1, \ldots, n$, were generated with mean function $\mu_{\mathcal{X}}(t) = t + \sin(t)$ and covariance function

constructed from the 10 first eigenfunctions of the Wiener Process with its correspondent eigenvalues. That is, for $0 \leq t \leq 1$, $\mathcal{X}_i(t) = \mu_{\mathcal{X}}(t) + \sum_{j=1}^{10} \rho_j \xi_{ij} \phi_j(t)$, where for $j \geq 1$, $\phi_j(t) = \sqrt{2}\sin((j-1/2)\pi t)$, $\rho_j = 1/((j-1/2)\pi)$ and the ξ_{ij} were generated from $N(0,1)$.

The function β_0 is defined as $\beta_0(t) = (t-0.25)^2\ \mathbf{1}_{[0.25,1]}$ and β_1 as

$$\beta_1(t) = \begin{cases} \frac{-2}{0.15^2}(t-0.45)^2 + 2 & \text{if } t \in [0.3, 0.6], \\ \frac{-1}{0.15^2}(t-0.85)^2 + 1 & \text{if } t \in [0.7, 1], \\ 0 & \text{otherwise.} \end{cases}$$

The noise ϵ_i is defined as follows, $\epsilon_i(t) = c_\epsilon \sum_{j=11}^{20} \rho_j \xi_{ij} \phi_j(t)$, where c_ϵ is a constant such that $SNR = 2$.

Results: The simulation results are presented in Figures 1, 2, and Table 1. The performance of the four estimators are illustrated.

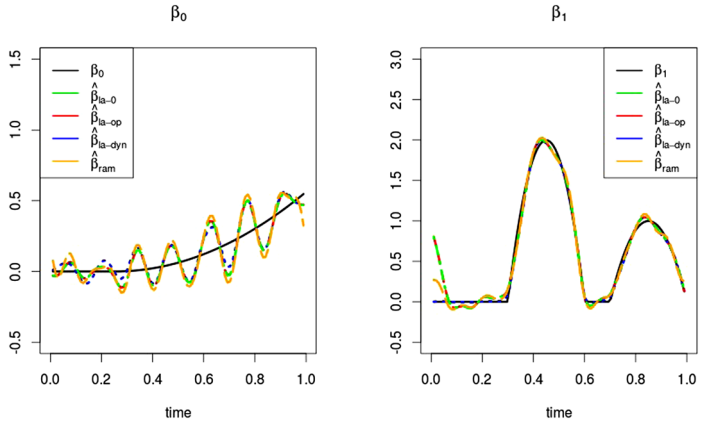

FIG 1. *An example of the estimation of β_0 and β_1 (solid black line) with a sample size $n = 70$. In green, the estimator without penalization ($\lambda = 0$); in red, the FREE with optimized parameter $\lambda > 0$; in blue, the FREE with the functional optimized parameter Λ; and in orange, Ramsay's estimator.*

We can see that, even under rather noisy conditions ($SNR = 2$), the estimators perform well. This shows their robustness. Furthermore, β_1 is better estimated than β_0 (see Figure 1) because of two reasons: (i) it is estimated before β_0 in (5.1) and (ii) since $\bar{\mathcal{X}}_n \approx \mu_{\mathcal{X}}$ has some periodicity, it introduces cycles on the estimators of β_0, which is monotone.

Lastly, let us remark that the FRRE computed with a functional regularization Λ_n gives in average better estimations. To understand better this fact, in Figure 3 we compare the mean of the 500 calibrated functional regularization parameters ($\bar{\Lambda}_n$) with the mean of the correspondent calibrated regularization parameters ($\bar{\lambda}_n$), which is equal to 0.5289 ($sd = 0.1096$).

The FRRE computed with a functional regularization Λ_n can reduce, if necessary, either the bias or the variance of the estimator in (3.1). This adaptability

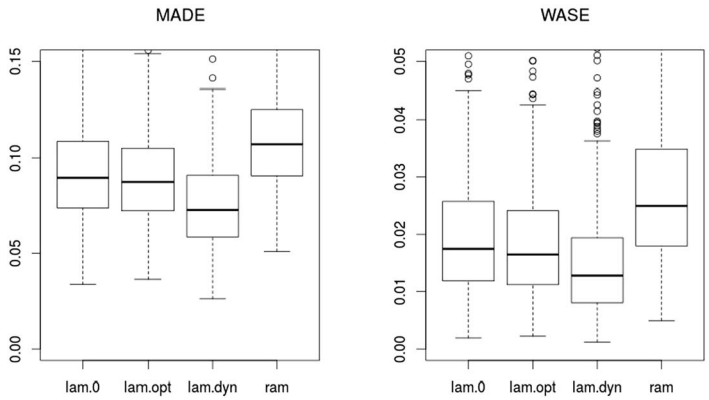

FIG 2. *Distribution of the evaluation criteria MADE (left panel) and WASE (right panel) for the four estimators over* 500 *simulated samples.*

TABLE 1
Means (and standard deviations) of the evaluation criteria MADE and WASE over 500 *simulated samples.*

	MADE	WASE
$\lambda = 0$	0.09171 (0.0252)	0.01985 (0.01129)
$\lambda_n > 0$	0.08973 (0.0238)	0.01873 (0.01047)
Λ_n	**0.07521 (0.0232)**	**0.01499 (0.0097)**
Ramsay	0.10938 (0.0255)	0.02784 (0.01386)

property makes it more efficient. An illustration is given in Figure 3. On the one hand, $\bar{\Lambda}_n$ penalizes much more in the intervals where β_1 is equal to zero to reduce the variance in (3.1). On the other hand, Λ_n is close to zero where $\beta_1 > 0$ to reduce the bias.

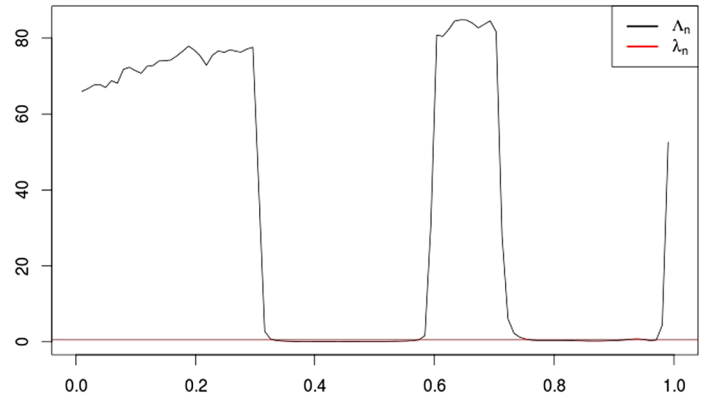

FIG 3. *The mean of the* 500 *calibrated functional regularization parameters* ($\bar{\Lambda}_n$) *and the mean of the correspondent calibrated regularization parameters* $\bar{\lambda}_n = 0.5289$.

5.1.2. Setting 2

We simulated samples with size $n = 100$. The input curves $\mathcal{X}_i$, for $i = 1, \ldots, n$, were generated with two white Gaussian noises. The first one over the subinterval $[0, 0.5]$ with a variance $\sigma^2_{X,I_1} = 0.5$, and the second one over $[0.5, 1]$ with a variance $\sigma^2_{X,I_2} = 0.5 * 1/10$. Accordingly, we have $\mathbb{E}[\mathcal{X}] = 0$ and the function $\mathbb{E}[|\mathcal{X}|^2]$ is constant over each of these subintervals.

Function β_0 is null and β_1 is defined as follows: $\beta_1 = 4(1-2\,t)^{1/2}\,\mathbf{1}_{[0,0.5]}$. The noise ϵ_i is defined in a similar way to $\mathcal{X}$, with a variance over $[0, 0.5]$ $\sigma^2_{\epsilon,I_1} = 0.3$ and over $[0.5, 1]$ $\sigma^2_{\epsilon,I_2} = 20 * 0.3$. Consequently, $SNR = 0.0873$ (very noisy situation).

Results: The simulation results are presented in Figures 4 and 5, and Table 2. The performance of the four estimators are illustrated.

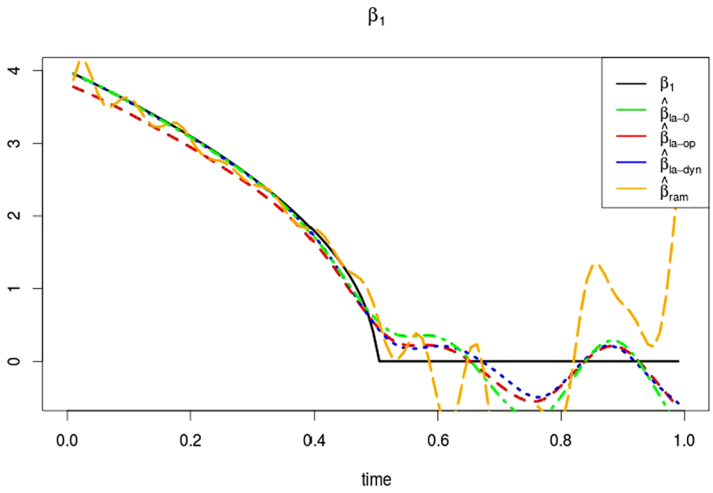

FIG 4. *An example of the estimation of β_1 (black solid line) with a sample size $n = 100$. In green, the estimator without penalization ($\lambda = 0$); in red, the FREE with optimized parameter $\lambda > 0$ optimal; in blue, the FREE with the functional optimized parameter Λ, and Ramsay's estimator in orange.*

Here, we see that under very noisy conditions ($SNR = 0.0873$), all the estimators perform well over the interval $[0, 0.5]$. Estimation performances differ over $[0.5, 1]$. The FRRE computed with a functional regularization Λ_n gives a more stable estimation.

Let us explain why the FRRE with a functional parameter performs better than the other estimators. First of all, in this setting we have $\mathbb{E}[\mathcal{X}] = 0$, then penalizing is needed to avoid dividing by zero when computing the FRRE. Thus, the estimator without penalization ($\lambda = 0$) is more unstable.

Secondly, the denominator of the FRRE ($1/n \sum_{i=1}^{n} |\mathcal{X}_i|^2$) behaves like $\mathbb{E}[|\mathcal{X}|^2]$, which is equal to 0.5 over $[0, 0.5]$ and to 0.05 over $[0.5, 1]$. Therefore, different

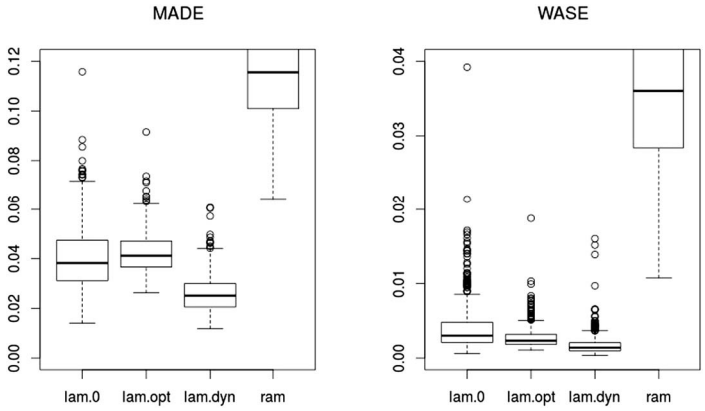

FIG 5. *Distribution of the evaluation criteria MADE (left panel) and WASE (right panel) for the four estimators over* 500 *simulated samples.*

TABLE 2
Means (and standard deviations) of the evaluation criteria MADE and WASE over 500 *simulated samples.*

	MADE	WASE
$\lambda = 0$	0.04049 (0.01324)	0.00404 (0.00342)
$\lambda_n > 0$	0.04258 (0.00836)	0.00274 (0.00151)
Λ_n	**0.02581 (0.00721)**	**0.00173 (0.00147)**
Ramsay	0.116 (0.02122)	0.03787 (0.01299)

penalization values are needed over each interval. A functional penalization like Λ is more flexible and consequently, it performs better than a constant penalization one.

Thirdly, given that the noise is 20 times bigger over $[0.5, 1]$ than over $[0, 0.5]$, a bigger penalization is needed over $[0.5, 1]$ to bound the variance in the bias-variance decomposition (3.1). This is better handled by the flexible functional penalization. Similarly, the bias is also better handled by a flexible penalization.

Finally, the FRRE estimators are more suitable than the estimator introduced by Ramsay et al (Ramsay et al. [14]). The main reason is that the FRRE estimators are pointwise defined, which avoids projecting the random functions onto a finite dimensional subspace that may be composed by too regular functions (Fourier basis). Thus, the approach we propose can better handle complex datasets of random functions such as realizations of the white Gaussian noise.

5.2. Dependence of the convergence rate and α

As stated in Theorem 3.5, the convergence rate of the estimator $(\|\hat{\beta}_n - \beta\|_{L^2})$ is bounded by $O_P\left(n^{-\gamma}\right)$, where $\gamma := \min\left[\frac{1}{2(2\alpha+1)}, \frac{1}{2} - \frac{1}{2(2\alpha+1)}\right]$. Therefore, this rate depends on α. In this way, the rate is directly related to the behavior of $\mathbb{E}[|X|^2(t)]$ around border points $(p \in C_{\beta,\partial X})$. This behavior is explained through

the polynomial lower bound function $|t-p|^{\alpha}$, according to hypothesis $A5$ (part a).

We present in setting 3 a case that explicitly shows the dependence of the convergence rate and α. In particular, we are interested in the behavior of $\|\hat{\beta}_n - \beta\|_{L^2}$ and of its upper bound, i.e.,

$$C_n := \frac{\lambda_n}{n} \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^{n} |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} + D_0 \frac{\sqrt{n}}{\lambda_n}, \tag{5.2}$$

where $D_0 = 10$ has been empirically chosen in order that (5.2) can be a bound of $\|\hat{\beta}_n - \beta\|_{L^2}$. From the proof of Theorem 3.5 (see section 8), we can see that C_n has a rate equal to $O_P\left(n^{-\gamma}\right)$.

To illustrate Theorem 3.5, we chose $p = 0 \in C_{\beta,\partial X}$ (see Assumption A5). The random functions X_i and Y_i are defined in a neighbourhood of p.

Setting 3

For each alpha value $\alpha \in \{0.001, 0.1, 0.25, 0.5, 1, 3, 9\}$, samples $(X_i, Y_i)_{i=1,\cdots,n}$, with sizes $n \in \{10^2, 10^3, 10^4, 5 \cdot 10^4, 10^5\}$, were simulated. More precisely, for each couple (n, α), $N = 50$ Monte Carlo runs were computed to obtain means of $\|\hat{\beta}_n - \beta\|_{L^2}$ and C_n.

The random functions X and Y are defined over the interval $[-1, 1]$. We considered the equispaced observation times, $[t_0, t_1, \cdots, t_{199}]$, with $t_k = -1 + 2k/199$ and $k = 0, \cdots, 199$.

The input functions X_i, for $i = 1, \cdots, n$, are realizations of

$$X(t) = |t|^{\alpha/2} + \frac{1}{40} \sum_{j=1}^{10} \rho_j \xi_{i,j} \phi_j(|t|),$$

where $t \in [-1, 1]$, and for $j \geq 1$, $\phi_j(|t|) = \sqrt{2}\sin((j - 1/2)\pi|t|)$, $\rho_j = 1/((j - 1/2)\pi)$ and the ξ_{ij} were generated from $N(0, 1)$. In this definition we use the ten first eigenfunctions of the Wiener Process with its corresponding eigenvalues. Similarly, the noise is defined as realizations of $\epsilon(t) = c_\epsilon \sum_{j=11}^{20} \rho_j \xi_{ij} \phi_j(t)$, where c_ϵ is a scalar, such that $SNR = 5$ (20 % of noise).

The functional coefficient is defined as $\beta(t) = 1.5 - t^2$. Lastly, the output functions Y_i are generated according to model (2.2).

From these definitions, $\mathbb{E}[|X(t)|^2] = |t|^{\alpha}$ and $p := 0 \in C_{\beta,\partial X}$.

Results: In Tables 3 and 4 we show the mean values of $\|\hat{\beta}_n - \beta\|_{L^2}$ and of its upper bound C_n, respectively.

Clearly, as the value of α increases, the convergence rate deteriorates due to the increasing bias. Specifically, when $\alpha >> 0$, $\mathbb{E}[|X|] \approx 0$ and then, the bias behaves like β in Equation (3.1) slowing down its rate.

The upper bound C_n behaves as expected for n large enough. That is, its convergence rate is very low when $\alpha \approx 0$, improves to reach its maximum value for $\alpha = 1/2$.

We can also see that $\|\hat{\beta}_n - \beta\|_{L^2}$ tends to 0 faster when α tends to 0. Indeed, for $\alpha \approx 0$, the function $\mathbb{E}[|X(t)|^2] = |t|^{\alpha} \approx 1$ over $[-1, -\delta] \cup [\delta, 1]$, where $\delta \in]0, 1[$

α	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 5 \cdot 10^4$	$n = 10^5$
0.001	**0.1557**	**0.0526**	**0.0170**	**0.0076**	**0.0054**
0.1	0.2466	0.1039	0.0414	0.0214	0.0161
0.25	0.4085	0.2210	0.1114	0.0673	0.0539
0.5	0.6882	0.4928	0.3351	0.2491	0.2179
1.0	1.0749	0.9538	0.8337	0.7529	0.7193
3.0	1.5200	1.4998	1.4788	1.4634	1.4566
9.0	1.6636	1.6618	1.6601	1.6589	1.6583

TABLE 3
Mean values of $\|\hat{\beta}_n - \beta\|_{L^2}$.

α	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 5 \cdot 10^4$	$n = 10^5$
0.001	10.1098	9.9839	9.9255	9.9002	9.8911
0.1	7.0595	5.7273	4.6830	4.0804	3.8472
0.25	5.0501	3.3832	2.2658	1.7148	1.5217
0.5	**3.8505**	**2.2710**	**1.3351**	**0.9179**	**0.7802**
1.0	3.2293	1.9538	1.2979	1.0244	0.9347
3.0	2.9095	2.0177	1.6718	1.5602	1.5286
9.0	2.7924	2.0411	1.7875	1.7183	1.7012

TABLE 4
Mean values of C_n.

is small. Using an equispaced grid around zero, we can assume that for all these observation times t_k, $|X(t_k)|^2 > 0.5$. Therefore, in Equation (3.1), we can bound the variance with $\frac{2}{n}\|\sum_{i=1}^{n} \epsilon_i X_i^*\|_{L^2} = O_P(1/\sqrt{n})$, and get an optimal rate for the variance. Similarly, we can show that the convergence rate of the bias $((O(\lambda_n/n))$ is high because when $\alpha \approx 0$, $\lambda_n/n \approx n^{-1/2}$ which is the parametric convergence rate.

In this way, we can see that when α tends to 0, both the variance and the bias have better convergence rates than $C_n = O_P\left(n^{-\gamma}\right)$. Thus the convergence rate of $\|\hat{\beta}_n - \beta\|_{L^2}$ reveals to be better than that of C_n, which is the upper bound obtained in Theorem 3.5. This bound is not optimal. The additional Proposition 8.5 in section 8 is stated to show how to improve the upper bound on compact sets.

6. Application

We illustrate the use of the estimators in (5.1) with the "gait data". These data have been processed by Ramsay et al. [14, p. 158] as an example of estimation in the FCM and can be found in the **R** package **fda**. The data "are measurements of angle at the hip and knee of 39 children as they walk through a single gait cycle. The cycle begins at the point where the child's heel under the leg being observed strikes the ground. For plotting simplicity we run time here over the interval $[0, 20]$, since there are 20 times at which the two angles are observed."

The main question the authors wanted to study was: "How much control does the hip angle have over the knee angle?". Accordingly, the hip angle curves are

the covariate $\mathcal{X}_i$ and the knee angle curves the response $\mathcal{Y}_i$. They model this interaction through the FCM with intercept (1.1).

The estimators of β_0 and β_1 (5.1) with optimized constant and functional parameters are presented in Figure 6. These estimators gave similar results as those obtained with **fda**, with a better computation time. Additionally, the empirical mean $\bar{\mathcal{Y}}_n$ is also compared to β_0 to see what happens if $\beta_1 = 0$, that is when the hip angle ($\mathcal{X}$) does not influence the knee angle ($\mathcal{Y}$). From Figure 6 (left panel) we see that a functional coefficient β_1 is required.

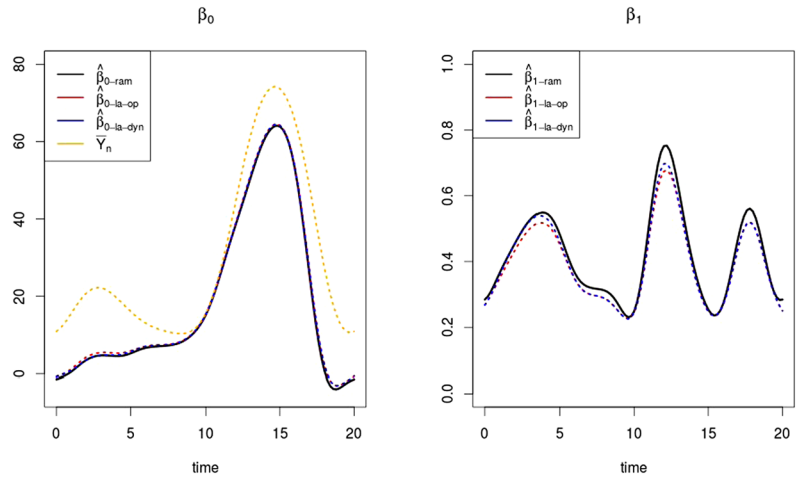

FIG 6. *The FRRE estimators of β_0 and β_1 (5.1) with optimized constant (in red) and functional (in blue) parameters compared to Ramsay's estimator (in black). The empirical mean $\bar{\mathcal{Y}}_n$ (in orange) is plotted in the first panel.*

7. Conclusions

In this paper we generalized the Ridge Regression method to define the FRRE estimator of the functional coefficient β_1 in the FCM (1.1). We proved its consistency for the L^2-norm, and obtained its rate of convergence over the whole real line, not only on compact sets.

From a practical point of view, we introduced two penalized estimators, one with a constant regularization parameter and the other with a functional one. The functional regularization is more flexible in case where the noise variance is changing over the estimation interval, or when the functional parameter β is close to 0. For both estimators, we provided a selection procedure through PCV.

In addition we compared this estimation method with that of Ramsay et al. [14, Ch. 10] in a simulation study and in an application. Both perform well under noisy conditions and in some cases the former is more robust, may better handle complex datasets of random functions and is faster to compute.

All these results open new perspectives for studying the FCM with several covariates and related models such as the convolution model (1.2), for which the properties of the Fourier transform allow to transpose the convergence results to an estimator based on the FRRE.

8. Proofs

8.1. Proof of Theorem 3.1

Let us first introduce a useful technical lemma. Here we will denote $\varphi := \mathbb{E}[|X|^2] \in C_0$.

Lemma 8.1. *Under hypotheses* $(A1)$ *and* $(A2)$ *of Theorem 3.1, if there exists a sequence of functions* $(f_n)_{n\geq 1} \subset C_0$ *such that* $\|f_n - \varphi\|_{C_0} \to 0$*, then there exist*

1. *a sequence* $(C_j)_{j\geq 1}$ *of subsets of* $\mathbb{R}$ *such that*

$$m\left(\limsup_{j\to+\infty} C_j\right) = m\left(\cap_{J\geq 1}[\cup_{j=1}^{J} C_j]\right) = 0,$$

   where m *is the Lebesgue measure,*
2. *a strictly increasing sequence of natural numbers* $(N_j)_{j\geq 1} \subset \mathbb{N}$ *and a sequence of real numbers* $(d_n)_{n\geq 1} \subset \mathbb{R}$*, with* $\lim_{n\to+\infty} d_n = 0$*,*

such that for every $j \geq 1$ *and* $n \in \{N_j, \cdots, N_{j+1}\}$*,*

$$\left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(\mathbb{R}\setminus C_j)} \leq d_n. \tag{8.1}$$

Proof of Lemma 8.1. To start the proof, we notice that $supp(\varphi) = supp(\mathbb{E}[|X|])$, hence $\overline{supp(|\beta|)} \subseteq \overline{supp(\varphi)}$ by hypothesis $(A1)$.

We define the sequence $\alpha_r := \sqrt{\frac{\lambda_r}{r}}$ which is decreasing to 0, and the sets $K_r^{\varphi} := \varphi^{-1}([\alpha_r, +\infty[)$ and $K_q^{\beta} := |\beta|^{-1}([1/q, +\infty[)$ for $r, q \in \mathbb{N}^+$. All these sets are compacts and cover the supports of φ and β respectively, that is $\cup_{r=1}^{\infty} \uparrow K_r^{\varphi} = supp(\varphi)$ and $\cup_{q=1}^{\infty} \uparrow K_q^{\beta} = supp(\beta)$.

Without loss of generality, we can suppose that there exists some $Q_1 \in \mathbb{N}$ such that $K_{Q_1}^{\beta} \neq \emptyset$ (otherwise $\beta \equiv 0$). Then we redefine for all $q \in \mathbb{N}$, $K_q^{\beta} := K_{Q_1+q}^{\beta}$.

Let us take a sequence δ_s decreasing to 0 and define for all $s \in \mathbb{N}$,

$$D_s := B_{\delta_s}(\partial supp(\varphi)) = \cup_{a\in\partial supp(\varphi)} B_{\delta_s}(a) \quad \text{and} \quad C_s := K_s^{\beta} \cap D_s,$$

with $B_{\delta_s}(a) :=]a - \delta_s, a + \delta_s[$. Clearly

$$K_1^{\beta} \setminus C_1 \subset int(supp(\varphi)) = supp(\varphi) = \cup_{r=1}^{\infty} K_r^{\varphi}.$$

since the supports of continuous functions are open.

Thus, from the definition of K_r^φ and the fact that α_r goes to zero, there exists $r_1 \in \mathbb{N}$ such that for all $r \geq r_1$, $K_1^\beta \setminus C_1 \subset K_r^\varphi$.

Moreover, from $(A2)$ there exists $\tilde{r}_1 > r_1$ such that, for all $r \geq \tilde{r}_1$,

$$\max_{r \geq \tilde{r}_1} \frac{\lambda_r}{r} \leq \frac{\lambda_{r_1}}{r_1}.$$

Considering $K_1^\beta \setminus C_1$, from the definition of $K_{r_1}^\varphi$ and the uniform convergence of $(f_n)_{n\geq 1}$ towards φ, we deduce that there exists $N_1 > \tilde{r}_1$ such that for all $n \geq N_1$ and $t \in K_{r_1}^\varphi$,

$$\frac{3}{4}\ \alpha_{r_1} \leq f_n(t) + \frac{\lambda_n}{n}.$$

Thus for all n such that $n \geq N_1$,

$$|\frac{\lambda_n}{n}|\left\|\frac{\beta}{f_n + \frac{\lambda_n}{n}}\right\|_{C_0(K_{r_1}^\varphi)} \leq |\frac{\lambda_n}{n}|\frac{4}{3\alpha_{r_1}}\|\beta\|_{C_0(\mathbb{R})} \leq \left(\max_{s\geq \tilde{r}_1}[|\frac{\lambda_s}{s}|]\right)\frac{4}{3\alpha_{r_1}}\|\beta\|_{C_0(\mathbb{R})}.$$

In particular we can deduce, for all $n \geq N_1 > r_1$,

$$|\frac{\lambda_n}{n}|\left\|\frac{\beta}{f_n + \frac{\lambda_n}{n}}\right\|_{C_0(K_1^\beta \setminus C_1)} \leq |\frac{\lambda_{r_1}}{r_1}|\frac{4}{3\alpha_{r_1}}\|\beta\|_{C_0(\mathbb{R})} \leq \sqrt{\frac{\lambda_{r_1}}{r_1}}\frac{4}{3}\|\beta\|_{C_0(\mathbb{R})}.$$

because of the definition of α_{r_1}.

Similarly

$$K_2^\beta \setminus C_2 \subset int(supp(\varphi)),$$

and there exists $r_2 > r_1$ such that for all $r \geq r_2$, $K_2^\beta \setminus C_{\delta_2} \subset K_r^\varphi$. From $(A2)$ there exists $\tilde{r}_2 > r_2$ such that $\max_{r\geq \tilde{r}_2} \frac{\lambda_r}{r} \leq \frac{\lambda_{r_2}}{r_2}$.

Again, given the definition of $K_{r_2}^\varphi$ and the uniform convergence of $(f_n)_{n\geq 1}$ towards φ, we deduce that there exists $N_2 > \tilde{r}_2$ such that for all $n \geq N_2$ and $t \in K_{r_2}^\varphi$,

$$\frac{3}{4}\ \alpha_{r_2} \leq f_n(t) + \frac{\lambda_n}{n}.$$

This yields that, for all n such that $n \geq N_2 > r_2$,

$$\left|\frac{\lambda_n}{n}\right|\left\|\frac{\beta}{f_n + \frac{\lambda_n}{n}}\right\|_{C_0(K_2^\beta \setminus C_2)} \leq \sqrt{\frac{\lambda_{r_2}}{r_2}}\frac{4}{3}\|\beta\|_{C_0(\mathbb{R})}.$$

We continue this way to build three strictly increasing sequences $r_j \uparrow \infty$, $\tilde{r}_j \uparrow \infty$ and $N_j \uparrow \infty$ such that for all $j \in \mathbb{N}$,

1. $N_j > \tilde{r}_j > r_j$,
2. $\forall r \geq r_j, \quad K_j^\beta \setminus C_j \subset K_r^\varphi$,
3. $\max_{r\geq \tilde{r}_j}[\frac{\lambda_r}{r}] \leq \frac{\lambda_{r_j}}{r_j}$,
4. $\forall n \geq N_j, |\frac{\lambda_n}{n}|\left\|\frac{\beta}{f_n+\frac{\lambda_n}{n}}\right\|_{C_0(K_j^\beta \setminus C_j)} \leq \sqrt{\frac{\lambda_{r_j}}{r_j}}\frac{4}{3}\|\beta\|_{C_0(\mathbb{R})}$.

Let n be an integer greater than N_1. Then there exists an integer j such that n belongs to the set $\{N_j, N_j+1, \cdots, N_{j+1}-1\}$. The following sequence (d_n) is then defined as follows:

$$d_n := \max\left\{\frac{4}{3}\sqrt{\frac{\lambda_{r_j}}{r_j}}\|\beta\|_{C_0(\mathbb{R})},\ \frac{1}{j}\right\}. \tag{8.2}$$

It is easy to see that this sequence goes to zero and from (8.2) we conclude that for all $n \in \{N_j, N_j+1, \cdots, N_{j+1}-1\}$,

$$|\frac{\lambda_n}{n}|\ \left\|\frac{\beta}{f_n+\frac{\lambda_n}{n}}\right\|_{C_0(\mathbb{R}\setminus C_j)} \leq d_n, \tag{8.3}$$

because of the definition of K_j^β (outside K_j^β, β is bounded by $1/j$) and the fact that $\mathbb{R}\setminus C_{\delta_j} = [K_j^\beta \setminus C_{\delta_j}] \cap [(K_j^\beta)^c \setminus C_{\delta_j}]$. □

Proof of Theorem 3.1. From the decomposition (3.1), we obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} \leq \left|\frac{\lambda_n}{n}\right| \left\|\frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}\right\|_{L^2} + \left\|\frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}\right\|_{L^2}.$$

Let us start by showing that

$$\left\|\frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}\right\|_{L^2} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right). \tag{8.4}$$

First we have

$$\mathbb{E}[\|\epsilon\, X^*\|_{L^2}^2] \leq \mathbb{E}[\|\epsilon\|_{C_0}^2]\ \mathbb{E}[\|X\|_{L^2}^2] < +\infty,$$

because of $(HA1_{FCM})$ and $(HA3_{FCM})$.

Now due to the moment monotonicity $\mathbb{E}[\|\epsilon\,\overline{X}\|_{L^2}] < +\infty$, $\epsilon\,\overline{X}$ is strongly integrable with the L^2-norm, so the function $\mathbb{E}[\epsilon\,\overline{X}]$ exists and belongs to L^2. From $(HA1_{FCM})$, $\mathbb{E}[\epsilon\,\overline{X}]$ is the zero function. We conclude that

$$\mathbb{E}[\epsilon\,\overline{X}] = 0 \qquad \text{and} \qquad \mathbb{E}[\|\epsilon\,\overline{X}\|_{L^2}^2] < +\infty,$$

which, from the CLT in L^2 (see Theorem 2.7 in Bosq [1, p. 51] and Ledoux and Talagrand [13, p. 276] for the rate of convergence), yields to

$$\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*\right\|_{L^2} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Finally (8.4) is obtained from the fact that

$$\left\|\frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |W_i|^2 + \frac{\lambda_n}{n}}\right\|_{L^2} \leq \left|\frac{n}{\lambda_n}\right| \left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*\right\|_{L^2} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right).$$

As $\frac{\sqrt{n}}{\lambda_n} \to 0$ by $(A3)$, we obtain the probability convergence of this part.

To conclude the proof, it is enough to show that

$$\left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} \xrightarrow{a.s.} 0. \tag{8.5}$$

To that purpose, we use the fact that

$$\left\| \frac{1}{n}\sum_{i=1}^n |X_i|^2 - \mathbb{E}[|X|^2] \right\|_{C_0} \xrightarrow{a.s.} 0,$$

which can be obtained by applying the Strong Law of Large Numbers (SLLN) (see Bosq [1, p. 47]) to the random function $|X|^2$. Notice here that $\mathbb{E}[|X|^2] \in C_0$.

Now for $\mathcal{S} := \{\omega \in \Omega : \|\frac{1}{n}\sum_i^n |X(\omega)|^2 - \varphi\|_{C_0} \to 0\}$, $P(\mathcal{S}) = 1$. Let us take an arbitrary and fixed value $\omega \in \mathcal{S}$. Then for $n \geq 1$ we define the sequence of functions $f_n := \frac{1}{n}\sum_{i=1}^n |X_i(\omega)|^2$. Clearly this sequence belongs to C_0 and $\|f_n - \varphi\|_{C_0} \to 0$. Thus we can use Lemma 8.1 which implies that there exists a sequence of subsets of $\mathbb{R}$, $(C_j)_{j\geq 1}$, a strictly increasing sequence of natural numbers $(N_j)_{j\geq 1} \subset \mathbb{N}$ and a sequence of real numbers $(d_n)_{n\geq 1} \subset \mathbb{R}$ converging to zero, such that inequality (8.1) holds.

At this point we define for $n \geq N_1$, $R_n := \frac{1}{d_n} \to \infty$ and the intervals $\bar{I}_n := [-R_n, +R_n]$. For $n \in \{N_j, N_j + 1, \cdots, N_{j+1} - 1\}$, by the triangular inequality and inequality (8.1),

$$\begin{aligned} |\tfrac{\lambda_n}{n}| \left\| \tfrac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\mathbb{R})} &\leq |\tfrac{\lambda_n}{n}| \left\| \tfrac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\bar{I}_n \cap C_j)} + |\tfrac{\lambda_n}{n}| \left\| \tfrac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\bar{I}_n \cap C_j^c)} + \\ &\quad + \|\beta\|_{L^2(\bar{I}_n^c)} \\ &\leq \|\beta\|_{L^2(C_j)} + |\tfrac{\lambda_n}{n}| \left\| \tfrac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(\mathbb{R}\setminus C_j)} \sqrt{2\,R_n} + \|\beta\|_{L^2(\bar{I}_n^c)}. \end{aligned}$$

In this way we obtain for every $n \in \{N_j, N_j + 1, \cdots, N_{j+1} - 1\}$,

$$|\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\mathbb{R})} \leq \|\beta\|_{L^2(C_j)} + d_n\sqrt{\frac{2}{d_n}} + \|\beta\|_{L^2(\bar{I}_n^c)}.$$

Thus

$$L := \lim_{n\to\infty} |\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\mathbb{R})} \leq \lim_{j\to\infty} \|\beta \cdot 1_{C_j}\|_{L^2(\mathbb{R})}.$$

Finally the sequence of functions $|\beta \cdot 1_{C_j}|$ is bounded by β and is pointwise convergent to zero almost everywhere because $\{t \in \mathbb{R} : \beta \cdot 1_{C_j}(t) \to 0\}^c \subset \cap_{l=1}^\infty \cup_{s\geq l} C_s \subset \cap_{l=1}^\infty \cup_{s\geq l} D_s \subset \cap_{l=1}^\infty D_l \subset \partial supp(\varphi)$ which is countable then with measure zero.

By the dominated convergence theorem, $\lim_{j\to\infty} \|\beta \cdot 1_{C_j}\|_{L^2} = 0$. Thus $L = 0$ and so (8.5) is proved because ω is an arbitrary element of $\mathcal{S}$ and $P(\mathcal{S}) = 1$. □

Proof of Proposition 3.4. For all independent realizations of X, we have $\mathbb{E}[\|X_n\|_{C_0([t_0-\delta,t_0+\delta])}] = 0$. Then for all $n \in \mathbb{N}$, the function X_n restricted to the interval $[t_0-\delta, t_0+\delta]$ is equal to zero almost surely. Thus over this interval $\hat{\beta}_n = 0$ (a.s.). If we define $C := \|\beta\|_{L^2([t_0-\delta,t_0+\delta])}$ we obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} \geq \|\hat{\beta}_n - \beta\|_{L^2([t_0-\delta,t_0+\delta])} = C \quad (a.s.). \qquad \square$$

8.2. Proof of Theorem 3.5

We use (3.1) and the triangle inequality to obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} \leq \left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(supp(|\beta|))} + \left\| \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2}.$$

The proof of

$$\left\| \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^1} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right)$$

is the same as in Theorem 3.1.

Hence, to finish the proof of Theorem 3.5, we have to show that

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(supp(|\beta|)\setminus J)} + \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(J)} = O_P(1), \tag{8.6}$$

which will lead to

$$\|\hat{\beta}_n - \beta\|_{L^2} = \left|\frac{\lambda_n}{n}\right| O_P(1) + O_P\left(\frac{\sqrt{n}}{\lambda_n}\right) = O_P\left(n^{-\gamma}\right).$$

The proof of (8.6) is based on the two following lemmas.

Lemma 8.2. *Under the assumptions of Theorem 3.5, we have*

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(supp(|\beta|)\setminus J)} = O_P(1).$$

Proof of Lemma 8.2. Throughout the proof, we use the following notations to simplify the writing. For all $n \geq 1$, $\bar{\lambda}_n := \frac{\lambda_n}{n}$, $S_n := \sum_{i=1}^n |X_i|^2$, $\bar{S}_n := \frac{S_n}{n}$, $A_n := |\beta|/(\bar{S}_n + \bar{\lambda}_n)$. The support of function $\varphi := \mathbb{E}[|X|^2]$ is $supp(\varphi) = supp(\mathbb{E}[|X|])$, so that $C_{\beta,\partial X} = supp(|\beta|) \setminus \partial(supp(\varphi))$. Finally, the set $C := supp(|\beta|) \setminus J$ satisfies $C \subset supp(\varphi)$.

Let us define for $j \geq 1$, $r_j := \|\varphi\|_{C_0}/2^j$, $r_0 := \|\varphi\|_{C_0} + 1$, the compact sets $K_0 := \emptyset$, $K_j := \varphi^{-1}([r_j, \infty[)$, and $D_j := K_j \setminus K_{j-1}$. So we have $\cup_{j\geq 1} \uparrow K_j = supp(\varphi)$ and we can cover $C = \cup_{j\geq 1}(C \cap D_j)$.

We obtain

$$\|A_n\|^2_{L^2(C)} = \sum_{j\geq 1} \|A_n\, 1_{\bar{S}_n\in[0,r_j/2]}\|^2_{L^2(C\cap D_j)} + \sum_{j\geq 1} \|A_n\, 1_{\bar{S}_n>r_j/2}\|^2_{L^2(C\cap D_j)}$$

$$\leq \tfrac{1}{\lambda_n^2} \sum_{j\geq 1} \|\beta\|^2_{C_0(C\cap D_j)}\, m(\bar{S}_n \in [0,r_j/2]\cap C\cap D_j) + \\ + \sum_{j\geq 1} \tfrac{2^2}{r_j^2}\, \tfrac{r_{j-1}^2}{r_{j-1}^2}\, \|\beta\|^2_{L^2(C\cap D_j)}.$$

Now for each $j\geq 1$, $\frac{r_{j-1}}{r_j} \leq \frac{r_0}{r_1}$ and in the set $C\cap D_j$, $\frac{\beta}{r_{j-1}} < \frac{\beta}{\varphi} \leq \frac{\beta}{r_j}$. Then $\|\beta\|_{C_0} \leq M_2 r_{j-1}$ because of part (c) of $(A5)$. Thus

$$\|A_n\|^2_{L^2(C)} \leq \tfrac{1}{\lambda_n^2} M_2^2\, (\tfrac{r_0}{r_1})^2 \sum_{j\geq 1} r_{j-1}^2\, m(\bar{S}_n \in [0,r_j/2]\cap C\cap D_j) + \\ + 4(\tfrac{r_0}{r_1})^2 \sum_{j\geq 1} \|\tfrac{\beta}{\varphi}\|^2_{L^2(C\cap D_j)}.$$

Moreover

$$\sum_{j\geq 1} \tfrac{r_j^2}{4}\, m(\bar{S}_n \in [0,r_j/2]\cap C\cap D_j) \leq \sum_{j\geq 1} \|(\varphi-\bar{S}_n)\, 1_{\bar{S}_n\in[0,r_j/2]}\|^2_{L^2(C\cap D_j)}$$

$$\leq \|\varphi-\bar{S}_n\|^2_{L^2(C)}.$$

Now we can bound A_n

$$\begin{aligned} \|A_n\|^2_{L^2(C)} &\leq \tfrac{1}{\lambda_n^2} M_2^2\, (\tfrac{r_0}{r_1})^2 \times 4\|\varphi-\bar{S}_n\|^2_{L^2(C)} + 4(\tfrac{r_0}{r_1})^2 \|\tfrac{\beta}{\varphi}\|^2_{L^2(C)} \\ &= 4M_2^2\, (\tfrac{r_0}{r_1})^2\, O_P((\tfrac{\sqrt{n}}{\lambda_n})^2) + 4(\tfrac{r_0}{r_1})^2 \|\tfrac{\beta}{\varphi}\|^2_{L^2(C)} = O_P(1). \end{aligned} \tag{8.7}$$

□

Lemma 8.3. *Under the assumptions of Theorem 3.5, we have*

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(J)} = O_P(1).$$

Proof of Lemma 8.3. We start the proof by considering the set $C_{\beta,\partial X}$. Since $supp(\varphi)$ is an open set in $\mathbb{R}$, it is a union of open intervals and $\partial(supp(\varphi))$ is countable. Besides, by hypothesis $(A5)$, for every $p\in C_{\beta,\partial X}$, there is an open neighborhood J_p, in which (a) holds. Thus for all $p\in C_{\beta,\partial X}$, $J_p\cap\partial(supp(\varphi)) = \{p\}$. These intervals J_p are countable and pairwise disjoint.

Now we suppose that $card(C_{\beta,\partial X}) = +\infty$ (the case where this set is finite is similar). We denote its elements as p_v, with $v\geq 1$. So J is the union of disjoint intervals $J = \cup_{v\geq 1} J_v$, where $J_v := J_{p_v}$, and part (b) of $(A5)$ can be written as $\sum_{v\geq 1} \|\beta\|^2_{C_0(J_v)} < M_1$.

Let us define $\xi_0 := \max\{\|\varphi\|_{C_0}, L_I^\alpha, \lambda_1^{2\alpha}+1\}$ and for $n\geq 1$, $\xi_n := \bar{\lambda}_n^{2\alpha}$. Clearly from $(A6)$, $\xi_n \downarrow 0$. We define for $l\geq 0$, the compact sets $K_l^\xi := \varphi^{-1}([\xi_l,\infty[)$, and $D_l^\xi := K_l^\xi \setminus K_{l-1}^\xi$. So we have $\cup\uparrow K_l^\xi = supp(\varphi)$ and we can cover $J_v\setminus\{p_v\} = \cup_{j\geq 1}(J_v\cap D_j^\xi)$ for each fixed $v\geq 1$. Moreover in D_l^ξ, $\frac{1}{\xi_{l-1}} < \frac{1}{\varphi} \leq \frac{1}{\xi_l}$.

Let us take a fixed $v \geq 1$. Given the fact that ξ_l is strictly decreasing to zero, by hypothesis $(A6)$, there exists a unique number $N_v \geq 1$ such that

$$\xi_{N_v} < \max_{t \in \partial(J_v)} |t - p_v|^\alpha \leq \xi_{N_v - 1}.$$

Then for every $n \geq N_v$,

$$\begin{aligned}
\|A_n\|^2_{L^2(J_v)} &= \textstyle\sum_{l=1}^n \|A_n\|^2_{L^2(J_v \cap D_l^\xi)} + \|A_n\|^2_{L^2(J_v \setminus K_n^\xi)} \\
&= \textstyle\sum_{l=1}^n \|A_n \, 1_{\bar{S}_n \in [0, \xi_l/2]}\|^2_{L^2(J_v \cap D_l^\xi)} + \\
&\quad + \textstyle\sum_{l=1}^n \|A_n \, 1_{\bar{S}_n \geq \xi_l/2}\|^2_{L^2(J_v \cap D_l^\xi)} \quad + \quad \|A_n\|^2_{L^2(J_v \setminus K_n^\xi)} \\
&\leq \|\beta\|^2_{C_0(J_v)} \left[\bar{\lambda}_n^{-2} \textstyle\sum_{l=1}^n m(\bar{S}_n \in [0, \xi_l/2] \cap J_v \cap D_l^\xi) \right] \\
&\quad + \quad \|\beta\|^2_{C_0(J_v)} \left[\textstyle\sum_{l=1}^n \frac{4}{\xi_l^2} m(J_v \cap D_l^\xi) + \quad \bar{\lambda}_n^{-2} m(J_v \setminus K_n^\xi) \right].
\end{aligned}$$

Using the inequality

$$\frac{\xi_n^2}{4} \sum_{l=1}^n m(\bar{S}_n \in [0, \xi_l/2] \cap J_v \cap D_l^\xi) \leq \|\varphi - \bar{S}_n\|_{L^2(J_v)},$$

we obtain

$$\begin{aligned}
\|A_n\|^2_{L^2(J_v)} \quad \leq \quad & \|\beta\|^2_{C_0(J_v)} \left[\bar{\lambda}_n^{-2} \tfrac{4}{\xi_n^2} \|\varphi - \bar{S}_n\|_{L^2(J_v)} + 4 \textstyle\sum_{l=1}^n \frac{\xi_{l-1}^2}{\xi_l^2} \frac{m(J_v \cap D_l^\xi)}{\xi_{l-1}^2} + \right. \\
& \left. + \quad \bar{\lambda}_n^{-2} m(J_v \setminus K_n^\xi) \right].
\end{aligned}$$

Because of $(A6)$, there exists $M_3 > 0$ such that for $l \geq 1$, $|\frac{\xi_{l-1}}{\xi_l}| \leq M_3$. Thus for $n \geq N_v$,

$$\begin{aligned}
\|A_n\|^2_{L^2(J_v)} \quad \leq \quad & \|\beta\|^2_{C_0(J_v)} \left[4 \, \bar{\lambda}_n^{-(2+4\alpha)} \|\varphi - \bar{S}_n\|_{L^2(J_v)} + \right. \\
& \left. + \quad 4 M_3^2 \|\tfrac{1}{\varphi}\|^2_{L^2(J_v \cap K_n^\xi)} + \bar{\lambda}_n^{-2} m(J_v \setminus K_n^\xi) \right].
\end{aligned}$$

Now for $t \in J_v \setminus K_n^\xi$ we have $0 \leq \varphi(t) < \xi_n$, then $|t - p_v|^\alpha \leq \varphi(t) < \xi_n$. In particular $J_v \setminus K_n^\xi \subset [p_v - \xi_n^{1/\alpha}, p_v + \xi_n^{1/\alpha}]$. Thus for $n \geq N_v$, $m(J_v \setminus K_n^\xi) \leq 2\xi_n^{1/\alpha} \leq 2\bar{\lambda}_n^2$.

Using these inequalities we can prove that for every $n < N_v$,

$$\|A_n\|^2_{L^2(J_v)} \leq \frac{1}{\bar{\lambda}_n^2} \|\beta\|^2_{L^2(J_v)},$$

and for $n \geq N_v$,

$$\|A_n\|^2_{L^2(J_v)} \leq 4 \|\beta\|^2_{C_0(J_v)} \left[n \|\bar{S}_n - \varphi\|^2_{L^2(J_v)} + M_3^2 \|\frac{1}{\varphi}\|^2_{L^2(J_v)} + 1/2 \right].$$

To finish the proof of this lemma, we bound the sequence $\|A_n\|^2_{L^2(J)} = \sum_{v\geq 1}\|A_n\|^2_{L^2(J_v)}$. In order to do this we define for each $n \geq 1$, the set $C_n := \{v \geq 1 : n < N_v\}$. We obtain

$$\begin{aligned}\|A_n\|^2_{L^2(J)} &\leq \bar{\lambda}_n^{-2}\|\beta\|^2_{L^2(\cup_{v\in C_n}J_v)} + \\ &+ \left(4\sum_{v\geq 1}\|\beta\|^2_{C_0(J_v)}\right)\left[n\|\bar{S}_n - \varphi\|^2_{L^2(J)} + M_3^2M_0^2 + 1/2\right] \\ &\leq \bar{\lambda}_n^{-2}\|\beta\|^2_{L^2(\cup_{v\in C_n}J_v)} + 4M_1\left[O_P(1) + M_3^2M_0^2 + 1/2\right].\end{aligned}$$

For each $n \geq 1$, $v \in C_n$ then $n < N_v$, hence $\xi_n \geq \max_{t\in\partial J_v}(t - p_v)^\alpha$, from what we deduce that $m(J_v) \leq 2\,\xi_n^{1/\alpha}$. We obtain for $n \geq 1$

$$\|\beta\|^2_{L^2(\cup_{v\in C_n}J_v)} \leq 2\xi_n^{1/\alpha}\sum_{v\in C_n}\|\beta\|^2_{C_0(J_v)} \leq 2\xi_n^{1/\alpha}\left[\sum_{v\geq 1}\|\beta\|^2_{C_0(J_v)}\right] = 2\xi_n^{1/\alpha}\left[M_1/4\right],$$

and thus for $n \geq 1$,

$$\begin{aligned}\|A_n\|^2_{L^2(J)} &\leq \bar{\lambda}_n^{-2}2\xi_n^{1/\alpha}\tfrac{M_1}{4} + 4M_1\left[O_P(1) + M_3^2M_0^2 + 1/2\right] \\ &\leq \tfrac{M_1}{2} + 4M_1\left[O_P(1) + M_3^2M_0^2 + 1/2\right] = O_P(1).\end{aligned}$$ □

Proof of Corollary 3.6. Direct computation using $\alpha < 1/2$ in Theorem 3.5. □

Proof of Corollary 3.9. As in the proof of Theorem 3.5, we only need to prove that

$$\left\|\frac{\beta}{\frac{1}{n}\sum_{i=1}^n|X_i|^2 + \frac{\lambda_n}{n}}\right\|^2_{L^2(supp(|\beta|))} = O_P(1).$$

To achieve this we use a similar method to that of Lemma 8.2. First note that hypothesis (*A4bis*) implies that, for all $t \in supp(\beta)$, $|\beta(t)|/\varphi(t)$ is finite. Consequently, $supp(\beta) \subset supp(\varphi)$.

Secondly we define $C := supp(\beta)$, $M_2 := \|\frac{\beta}{\mathbb{E}[|X|^2]}\|_{L^\infty}$, $r_0 := \|\varphi\|_{C_0} + 1$ and, for $j \geq 1$, $r_j := \|\varphi\|_{C_0}/2^j$. Then we apply the same method as in Lemma 8.2. This leads to the inequality (8.7) which implies what we wanted. □

Proof of Theorem 3.10. We start with the decomposition

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} = \left|\frac{\lambda_n}{n}\right|\left\|\frac{\beta}{\frac{1}{n}\sum_{i=1}^n|X_i|^2 + \frac{\lambda_n}{n}}\right\|_{L^2(K)} + \left\|\frac{\frac{1}{n}\sum_{i=1}^n\epsilon_iX_i^*}{\frac{1}{n}\sum_{i=1}^n|X_i|^2 + \frac{\lambda_n}{n}}\right\|_{L^2(K)}.$$

The proof of $\left\|\frac{\frac{1}{n}\sum_{i=1}^n\epsilon_iX_i^*}{\frac{1}{n}\sum_{i=1}^n|X_i|^2+\frac{\lambda_n}{n}}\right\|_{L^2(K)} = O_P(\frac{\sqrt{n}}{\lambda_n})$ is the same as in Theorem 3.1. We finish the proof of the theorem by showing that

$$\left\|\frac{\beta}{\frac{1}{n}\sum_{i=1}^n|X_i|^2 + \frac{\lambda_n}{n}}\right\|_{L^2(K)} = O_P(1). \tag{8.8}$$

Given that $K \subset supp(\varphi)$, there exists a positive number $s_1 > 0$ such that $K \subset K_{s_1}^{\varphi}$, where $K_{s_1}^{\varphi} := \varphi^{-1}([s_1, \infty[)$ is a compact in $\mathbb{R}$. We define $s := s_1/2$. We have for every $n \in \mathbb{N}$,

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} \right\|_{L^2(K)} \leq \left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} + \left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [s,\infty[} \right\|_{L^2(K)}.$$

Clearly, the first part above is bounded by

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [s,\infty[} \right\|_{L^2(K)} \leq \frac{1}{s} \ \|\beta\|_{L^2(K)} = O_P(1).$$

The second part is bounded as follows

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} \leq \frac{1}{\bar{\lambda}_n} \left\| \beta 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} \leq \frac{\|\beta\|_{C_0}}{\bar{\lambda}_n} \sqrt{m(K \cap \bar{S}_n \in [0,s])}.$$

Moreover, thanks to hypothesis $(A3)$, we have $\|\bar{S}_n - \varphi\|_{L^2(K)} = O_P(\frac{1}{\sqrt{n}})$. This inequality, together with the fact that $|\bar{S}_n - \varphi| > s$ whenever $\bar{S}_n \in [0, s]$, allows to obtain

$$\begin{aligned} \|\bar{S}_n - \varphi\|_{L^2(K)} &\geq \|(\bar{S}_n - \varphi) 1_{\bar{S}_n \in [0,s]}\|_{L^2(K)} \geq \sqrt{\textstyle\int_K |s|^2 1_{\bar{S}_n \in [0,s]} dm} \\ &\geq |s| \ \sqrt{m(K \cap \bar{S}_n \in [0,s])}. \end{aligned}$$

In this way, $\sqrt{m(K \cap \bar{S}_n \in [0,s])} = O_P(\frac{1}{\sqrt{n}})$ and as a consequence

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} \leq \frac{\|\beta\|_{C_0}}{\bar{\lambda}_n} \ O_P(\frac{1}{\sqrt{n}}) = O_P(\frac{\sqrt{n}}{\lambda_n}),$$

which finishes the proof of (8.8). □

8.3. Further results on the convergence rates

Corollary 8.4. *Under the hypotheses* $(A2)$, $(A3)$ *and* *(A1bis)* $\overline{supp(\beta)}$ *is compact and* $\overline{supp(|\beta|)} \subset supp(\mathbb{E}[|X|])$, *we obtain*

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n} \right] \right).$$

Proof of Corollary 8.4. Take $K = \overline{supp(\beta)}$ in Theorem 3.10 to upper bound $\|\hat{\beta}_n - \beta\|_{L^2(K)}$. Finally, we have $\|\hat{\beta}_n - \beta\|_{L^2(K^c)} \leq O_P(\frac{\sqrt{n}}{\lambda_n})$ because, first $\beta \equiv 0$ over K^c, which implies that the bias is null outside K (see (3.1)), and secondly, $O_P(\frac{\sqrt{n}}{\lambda_n})$ is the natural upper bound of the variance over K^c. Thus, using the bias-variance decomposition we upper bound $\|\hat{\beta}_n - \beta\|_{L^2(\mathbb{R})}$ as we wanted. □

Under more restricted hypotheses we can obtain the **optimal** rate of convergence. This is shown in Proposition 8.5 for the model (2.2).

Proposition 8.5. *Under the hypotheses* $(A1bis)$, $(A2)$, $(A3)$ *and*

(A4ter) There is $m_X > 0$ *s.t.* $|X| > m_X$ *almost surely over* $\overline{supp(|\beta|)}$,

we obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\frac{\lambda_n}{n}\right).$$

Furthermore, under the hypotheses $(A1bis)$, $(A3)$, $(A4ter)$ *and by replacing* $(A2)$ *with*

(A2bis) $(\lambda_n)_{n\geq 1} \subset \mathbb{R}^+$ *is the constant sequence equal to* $\lambda > 0$,

we obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Proof of Proposition 8.5. We start with the decomposition

$$\|\hat{\beta}_n - \beta\|_{L^2} \leq \|\hat{\beta}_n - \beta\|_{L^2(K)} + \|\hat{\beta}_n - \beta\|_{L^2(K^c)},$$

where $K := \overline{supp(\beta)}$.

First, we obtain the convergence rates over K considering the bias variance decomposition

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} = \left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)} + \left\| \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)}.$$

We deduce from Hypothesis $(A4ter)$ that almost surely

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} \leq \left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{m_X} \right\|_{L^2(K)} + \left\| \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{m_X} \right\|_{L^2(K)},$$

which, under hypothesis $(A2)$, implies

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} = O_P(\frac{\lambda_n}{n}) + O_P(\frac{1}{\sqrt{n}}) = O_P(\frac{\lambda_n}{n}).$$

Likewise, given that the bias is null over K^c, we obtain from a similar bias-variance decomposition over K^c that

$$\|\hat{\beta}_n - \beta\|_{L^2(K^c)} \leq 0 + \left\| \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{m_X} \right\|_{L^2(K)} = O_P(\frac{1}{\sqrt{n}}).$$

Thus, we get the convergence rate of $\|\hat{\beta}_n - \beta\|_{L^2}$ by adding the rates over K and K^c, that is, $O_P(\frac{\lambda_n}{n})$, under hypothesis $(A2)$.

Finally, when hypothesis $(A2bis)$ holds, the upper bound of $\|\hat{\beta}_n - \beta\|_{L^2(K)}$ is $O_P(\frac{\lambda}{n}) + O_P(\frac{1}{\sqrt{n}})$, with $\lambda > 0$ constant. This bound is equal to $O_P(\frac{1}{\sqrt{n}})$. The bound of $\|\hat{\beta}_n - \beta\|_{L^2(K^c)}$ is $O_P(\frac{1}{\sqrt{n}})$. Then, by adding both we get the optimal convergence rate. □

8.4. Proof of Theorem 3.11

From the decomposition (3.1) we obtain

$$\mathbb{E}[\|\hat{\beta}_n - \beta\|_{L^2}^2] \leq 2|\bar{\lambda}_n|^2 \mathbb{E}\|\frac{\beta}{\bar{S}_n + \bar{\lambda}_n}\|_{L^2}^2 + \frac{2}{|\bar{\lambda}_n|^2}\mathbb{E}\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i X_i^*\|_{L^2}^2,$$

where $\bar{\lambda}_n := \frac{\lambda_n}{n}$ and $\bar{S}_n := \frac{\sum_{i=1}^n |X_i|^2}{n}$.

Thus to finish this proof we need to prove two things:

$$\mathbb{E}\left\|\frac{\beta}{\bar{S}_n + \bar{\lambda}_n}\right\|_{L^2}^2 = O(1) \qquad \text{and} \qquad \mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i X_i^*\right\|_{L^2}^2 = O(\frac{1}{n}).$$

Let us prove the first equality. We know that hypothesis ($A4bis$) implies that the set $C_{\beta,\partial X} := supp(|\beta|)\backslash\partial(supp(\varphi))$ is empty (see proof of the Corollary 3.9). For this reason, by taking $J := \emptyset$ the hypotheses ($A4$) and ($A5$) in Theorem 3.5 will hold.

Now we can extend the inequality (8.7) of Lemma 8.2 to the whole real line because $C = supp(\beta)$ in this inequality. Then we have

$$\left\|\frac{\beta}{\bar{S}_n + \bar{\lambda}_n}\right\|_{L^2}^2 \leq \frac{1}{\bar{\lambda}_n^2} M_2^2 M_3 \|\varphi - \bar{S}_n\|_{L^2}^2 + M_3\|\frac{\beta}{\varphi}\mathbf{1}_{supp(\beta)}\|_{L^2}^2,$$

where $M_3 := 16(\frac{\|\varphi\|_{C_0}+1}{\|\varphi\|_{C_0}})^2$. The second term in the right side of this inequality is non random. Then we need to prove that the expectation of the first term in this side goes to zero, that is

$$\frac{1}{\bar{\lambda}_n^2} M_2^2 M_3 \mathbb{E}[\|\varphi - \bar{S}_n\|_{L^2}^2] \to 0.$$

To prove this, let us recall that $\frac{n}{\bar{\lambda}_n^2} \to 0$ by hypothesis ($A2$). So we only need to show that there exists a constant $d > 0$ such that for all $n \in \mathbb{N}$, $\mathbb{E}[\|\varphi - \bar{S}_n\|_{L^2}^2] \leq \frac{d}{n}$.

From hypothesis ($A3$) we can prove that $|\varphi - \bar{S}_n|^2$ is a random function belonging to $L^1(\mathbb{R}, \mathbb{R})$ and that $\mathbb{E}\int_{\mathbb{R}} |\varphi - \bar{S}_n|^2 < \infty$. Thus by Fubini and Tonelli Theorems (see Brezis [2, p. 91]) and thanks to the independence of the X_i we have

$$\begin{aligned}
\mathbb{E}[\|\varphi - \bar{S}_n\|_{L^2}^2] &= \int_{\mathbb{R}} \mathbb{E}[|\varphi - \tfrac{1}{n}\textstyle\sum_{i=1}^n |X_i|^2|^2] \\
&\leq \tfrac{1}{n}\int_{\mathbb{R}} \mathbb{E}[|\varphi - |X|^2|^2] \\
&\leq \tfrac{2}{n}\left\{\int_{\mathbb{R}} \mathbb{E}[|\varphi|^2] + \int_{\mathbb{R}} \mathbb{E}[|X|^4]\right\}
\end{aligned}$$

Now again by hypothesis ($A3$) we get

$$\mathbb{E}[\|\varphi - \bar{S}_n\|_{L^2}^2] \leq \frac{2}{n}\left\{\|\varphi\|_{L^2}^2 + \mathbb{E}\||X|^2\|_{L^2}^2\right\}.$$

Thus by putting $d := 2\|\varphi\|_{L^2}^2 + 2\mathbb{E}\||X|^2\|_{L^2}^2 < \infty$ we obtain the first equality.

Next, to finish this proof we will prove the second equality, namely

$$\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i X_i^*\right\|_{L^2}^2 = O(\frac{1}{n}).$$

From the hypotheses $(HA1_{FCM})$ and $(HA3_{FCM})$ it can be proved that the random function $|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i X_i^*|^2$ belongs to $L^1(\mathbb{R},\mathbb{R})$ and that its expectation is upper bounded. Thus thanks to the independence of the ϵ_i and X_i we obtain what we wanted

$$\begin{aligned}\mathbb{E}[\|\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\epsilon_i X_i^*\|_{L^2}^2] &= \textstyle\int_{\mathbb{R}}\mathbb{E}[|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i X_i^*|^2]\\ &= \tfrac{1}{n}\textstyle\int_{\mathbb{R}}\mathbb{E}[|\epsilon X^*|^2]\\ &\leq \tfrac{1}{n}\mathbb{E}\|X\|_{C_0}^2\,\mathbb{E}\|\epsilon\|_{L^2}^2 = O(\tfrac{1}{n}).\end{aligned}$$

Proof of Proposition 3.12. The conditional expectation comes directly from the decomposition (3.1). To compute the variance let us define for $i = 1, \cdots, n$, the random functions $g_i := \frac{X_i^*}{\frac{1}{n}\sum_{i=1}^{n}|X_i|^2+\frac{\lambda_n}{n}}$. Since the g_i are independent of the ϵ_i we obtain

$$\begin{aligned}\mathrm{Var}[\hat{\beta}_n \,|\, X_1, \cdots, X_n] &= \mathbb{E}[|\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\epsilon_i g_i|^2]\\ &= \tfrac{1}{n^2}\textstyle\sum_{i=1}^{n}[\mathbb{E}(|\epsilon_i|^2)]|g_i|^2\\ &= \tfrac{1}{n^2}\mathbb{E}(|\epsilon|^2)]\textstyle\sum_{i=1}^{n}[|g_i|^2]\\ &= \frac{\mathbb{E}(|\epsilon|^2)}{n}\left[\frac{\frac{1}{n}\sum_{i=1}^{n}|X_i|^2}{(\frac{1}{n}\sum_{i=1}^{n}|X_i|^2+\frac{\lambda_n}{n})2}\right]\\ &= \frac{\mathbb{E}(|\epsilon|^2)}{\sum_{i=1}^{n}|X_i|^2}\,D_X^2,\end{aligned}$$

where D_X is a function defined as follows $D_X := \frac{\frac{1}{n}\sum_{i=1}^{n}|X_i|^2}{\frac{1}{n}\sum_{i=1}^{n}|X_i|^2+\frac{\lambda_n}{n}}$. Here we need $D_X(t) > 0$, otherwise $X_1(t) = \cdots = X_n(t) = 0$ and nothing can be inferred about $\beta(t)$.

Next, let $t \in \mathbb{R}$ be fixed and such that $\epsilon(t) \sim N(0, \sigma_\epsilon^2)$ and $(\epsilon_i(t))_{i=1,\cdots,n}$ is an i.i.d sample. To simplify the proof let us define for $i = 1, \cdots, n$, $x_i := X_i(t) \in \mathbb{R}$, $y_i := Y_i(t) \in \mathbb{R}$, $\varepsilon_i := \epsilon_i(t) \in \mathbb{R}$ and $b_1 := \beta(t) \in \mathbb{R}$. Thanks to these conditions the set $(x_i, y_i)_{i=1,\cdots,n}$ is an i.i.d sample of the linear regression model, $y_i = b_1 x_i + \varepsilon_i$. For this model the OLS estimator of b_1 is $\hat{b}_1 := \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n}|x_i|^2}$.

The OLS estimator (see Cornillon and Matzner-Lober [3, p. 12]) is unbiased ($\mathbb{E}[\hat{b}_1] = b_1$), its variance is $\mathrm{Var}(\hat{b}_1) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}|x_i|^2}$ and it follows a normal law. Furthermore (see Cornillon and Matzner-Lober [3, p. 49]),

$$\frac{\hat{b}_1 - b_1}{\tilde{\sigma}_\epsilon(\sum_{i=1}^{n}|x_i|^2)^{-1/2}} \sim \mathcal{T}(n-1),$$

where $\tilde{\sigma}_\epsilon := \frac{1}{n-1}\sum_{i=1}^n |y_i - \hat{b}_1 x_i|^2$ is an unbiased estimator of σ_ϵ^2.

From these properties and the fact that $\hat{b}_1 = \frac{\hat{\beta}_n(t)}{D_X(t)}$ we obtain: i) $\hat{\sigma}_\epsilon = \tilde{\sigma}_\epsilon$ is unbiased, ii)

$$\frac{\hat{\beta}_n(t) - \beta(t) D_X(t)}{\hat{\sigma}_\epsilon D_X(\sum_{i=1}^n |X_i(t)|^2)^{-1/2}} = \frac{\hat{b}_1 - b_1}{\tilde{\sigma}_\epsilon(\sum_{i=1}^n |x_i|^2)^{-1/2}} \sim \mathcal{T}(n-1),$$

and iii) the confidence interval of $\beta(t) = b_1$. □

8.5. Proofs of the results of Section 4

Proof of Proposition 4.1. We only need to prove that for every $i = 1, \cdots, n$,

$$Y_i - \hat{\beta}_n^{(-i)} X_i = \frac{Y_i - \hat{\beta}_n X_i}{1 - A_{i,i}}. \tag{8.9}$$

Let us take an arbitrary $i \in \{1, \cdots, n\}$. We define for each $j = 1, \cdots, n$,

$$\tilde{Y}_j := \begin{cases} Y_j & \text{if } j \neq i, \\ \hat{\beta}_n^{(-i)} X_j & \text{otherwise.} \end{cases}$$

Because $\hat{\beta}_n^{(-i)} = \frac{\sum_{l\neq i}^n Y_l X_l^*}{\sum_{l\neq i}^n |X_l|^2 + \lambda_n}$ by definition, we have

$$\begin{aligned} \frac{\sum_{l=1}^n \tilde{Y}_l X_l^*}{S_n + \lambda_n} &= \frac{\sum_{l\neq i}^n Y_l X_l^*}{S_n + \lambda_n} + \frac{\hat{\beta}_n^{(-i)} |X_i|^2}{S_n + \lambda_n} \\ &= \hat{\beta}_n^{(-i)} \left[\frac{\sum_{l\neq i}^n |X_l|^2 + \lambda_n}{S_n + \lambda_n} + \frac{|X_i|^2}{S_n + \lambda_n} \right] = \hat{\beta}_n^{(-i)}. \end{aligned}$$

Then

$$\hat{\beta}_n X_i - \hat{\beta}_n^{(-i)} X_i = \frac{\sum_{l=1}^n Y_l X_l^* - \sum_{l=1}^n \tilde{Y}_l X_l^*}{S_n + \lambda_n} X_i = \frac{Y_i - \hat{\beta}_n^{(-i)} X_i}{S_n + \lambda_n} |X_i|^2,$$

from what we obtain

$$1 - \frac{Y_i - \hat{\beta}_n X_i}{Y_i - \hat{\beta}_n^{(-i)} X_i} = \frac{\hat{\beta}_n X_i - \hat{\beta}_n^{(-i)} X_i}{Y_i - \hat{\beta}_n^{(-i)} X_i} = \frac{|X_i|^2}{S_n + \lambda_n} = A_{i,i},$$

which implies (8.9). □

Proof of Proposition 4.3. It is similar to that of Proposition 4.1. □

Proof of Theorem 4.4. We use (3.1), the triangle inequality and the hypothesis (*A2bis*) to obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} \leq \left| \frac{b\, m_{\Lambda_n}}{n} \right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{m_{\Lambda_n}}{n}} \right\|_{L^2(supp(|\beta|))} + \left\| \frac{\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{m_{\Lambda_n}}{n}} \right\|_{L^2}.$$

The proof of

$$\left\| \frac{\frac{1}{n}\sum_{i=1}^{n} \epsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^{n} |X_i|^2 + \frac{m_{\Lambda_n}}{n}} \right\|_{L^1} = O_P\left(\frac{\sqrt{n}}{m_{\Lambda_n}}\right),$$

is the same as in Theorem 3.1.

Thus we only need to prove

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^{n} |X_i|^2 + \frac{m_{\Lambda_n}}{n}} \right\|_{L^2(supp(|\beta|))} = O_P(1),$$

since $b > 0$ is constant and does not modify the rate of convergence. To prove this equality we use the same techniques as in the proof of Corollary 3.9 with $\lambda_n := m_{\Lambda_n}$. □

Acknowledgements

We are grateful to the associate editor and the anonymous referees for valuable comments that helped us to improve the paper.

References

[1] Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, volume 149 of *Lectures Notes in Statistics*. Springer-Verlag, New York. MR1783138

[2] Brezis, H. (2010). *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media. MR2759829

[3] Cornillon, P.-A. and Matzner-Lober, E. (2010). *Régression avec R*. Springer Science & Business Media.

[4] Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the r package fda. usc. *Journal of Statistical Software*, 51(4):1–28.

[5] Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall / CRC Press.

[6] Hall, P., Horowitz, J. L., et al. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91. MR2332269

[7] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):109–126.

[8] Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58(3):54–59.

[9] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

[10] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, volume 200 of *Springer Series in Statistics*. Springer, New York. MR2920735

[11] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. MR3379106

[12] Huh, M.-H. and Olkin, I. (1995). Asymptotic aspects of ordinary ridge regression. *American Journal of Mathematical and Management Sciences*, 15(3–4):239–254.

[13] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces, Isoperimetry and Processes*, volume 23 of *A Series of Modern Surveys in Mathematics Series*. Springer-Verlag, Berlin.

[14] Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.

[15] Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer, New York, second edition.

[16] Şentürk, D. and Müller, H.-G. (2010). Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association*, 105(491):1256–1264.

[17] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, 10(4):1040–1053. MR0673642

[18] Wu, C. O., Chiang, C.-T., and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American statistical Association*, 93(444):1388–1402. MR1666635

[19] Yao, F., Müller, H.-G., and Jane-Ling, W. (2005a). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903.

[20] Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.