TECHNICAL REPORT

11042

On the effect of noisy observations of the regressor in a functional linear model

BERESWILL, M. and J. JOHANNES

IAP STATISTICS NETWORK

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Second International Workshop on
Functional and Operatorial Statistics.
Santander, June 16-18, 2011

On the effect of noisy observations of the regressor in a functional linear model

Mareike Bereswill and Jan Johannes

Ruprecht-Karls Universität Heidelberg, IAM, D-69120 Heidelberg, Germany
mareike.bereswill@web.de
Université catholique de Louvain, ISBA, B-1348 Louvain-la-Neuve, Belgium
jan.johannes@uclouvain.be

Abstract

We consider the estimation of the slope function in functional linear regression, where a scalar response Y is modeled in dependence of a random function X, when Y and only a panel $Z_1, \dots, Z_L$ of noisy observations of X are observable. Assuming an iid. sample of $(Y, Z_1, \dots, Z_L)$ we derive in terms of both, the sample size and the panel size, a lower bound of a maximal weigthed risk over certain ellipsoids of slope functions. We prove that a thresholded projection estimator can attain the lower bound up to a constant.

This work was supported by the IAP research network no. P6/03 of the Belgian Government (Belgian Science Policy).

1. Introduction

A common problem in a diverse range of disciplines is the investigation of the dependence of a real random variable Y on the variation of an explanatory random function X (see for instance Ramsay and Silverman [2005] and Ferraty and Vieu [2006]). We assume that X takes its values in an infinite dimensional separable Hilbert space $\mathbb{H}$ which is endowed with an inner product $\langle \cdot, \cdot \rangle$ and its associated norm $\|\cdot\|$. In functional linear regression the dependence of the response Y on the regressor X is then modeled by

$$Y = \langle \beta, X \rangle + \sigma\, \varepsilon, \quad \sigma > 0, \tag{1a}$$

where $\beta \in \mathbb{H}$ is unknown and the error ε has mean zero and variance one. In this paper we suppose that we have only access to Y and a panel of noisy observations of X,

$$Z_\ell = X + \varsigma\, \Xi_\ell, \quad \varsigma \geqslant 0, \quad \ell = 1, \dots, L, \tag{1b}$$

where $\Xi_1, \dots, \Xi_L$ are measurement errors. One objective is then the non-parametric estimation of the slope function β based on an iid. sample of $(Y, Z_1, \dots, Z_L)$.

In recent years the non-parametric estimation of the slope function β from a sample of (Y, X) has been of growing interest in the literature (c.f. Cardot et al. [1999], Marx and Eilers [1999], Bosq [2000] or Cardot et al. [2007]). In this paper we follow an approach based on dimension reduction and thresholding techniques, which has been proposed by Cardot and Johannes [2010] and borrows ideas from the inverse problems community (c.f. Efromovich and Koltchinskii [2001] and Hoffmann and Reiß [2008]).

The objective of this paper is to establish a minimax theory for the non-parametric estimation of β in terms of both, the size L of the panel $Z_1, \dots, Z_L$ of noisy measurements of X and the size n of the sample of $(Y, Z_1, \dots, Z_L)$. In order to make things more formal let us reconsider model (1a) - (1b). Given an orthonormal basis $\{\psi_j\}_{j\geqslant 1}$ in $\mathbb{H}$ (not necessarily corresponding to the eigenfunctions of Γ) we assume real valued random variables $\xi_{j,\ell} := \langle \Xi_\ell, \psi_j \rangle$ and observable blurred versions of the coefficient $\langle X, \psi_j \rangle$ of X,

$$Z_{j,\ell} := \langle X, \psi_j \rangle + \varsigma\, \xi_{j,\ell}, \quad \ell = 1, \dots, L \text{ and } j \in \mathbb{N}. \tag{2}$$

The motivating example for our abstract framework consists in irregular and sparse repeated measures of a contaminated trajectory of a random function $X \in L^2[0,1]$ (c.f. Yao et al. [2005] and references therein). To be more precise, suppose that there are L uniformly-distributed and independent random measurement times $U_1, \dots, U_L$ for X. Let $V_\ell = X(U_\ell) + \eta_\ell$ denote the observation of the random trajectory X at a random time U_ℓ contaminated with measurement error η_ℓ, $1 \leqslant \ell \leqslant L$. The errors η_ℓ are assumed to be iid. with mean zero and finite variance. If the random function X, the random times $\{U_\ell\}$ and the errors $\{\eta_\ell\}$ are independent, then, it is easily seen that for each $\ell = 1, \dots, L$ and $j \in \mathbb{N}$ the observable quantity $Z_{j,\ell} := V_\ell \psi_j(U_\ell)$ is just a blurred version of the coefficient $\langle X, \psi_j \rangle$ corrupted by an uncorrelated additive measurement error $V_\ell \psi_j(U_\ell) - \langle X, \psi_j \rangle$. Moreover, those errors are uncorrelated for all $j \in \mathbb{N}$ and different values of ℓ. It is interesting to note that recently Crambes et al. [2009] prove minimax-optimality of a spline based estimator in the situation of deterministic measurement times. However, the obtained optimal rates are the same as for a known regressor X since the authors suppose the deterministic design to be sufficiently dense. In contrast to this result we seek a minimax theory covering also sparse measurements. In particular, it enables us to quantify the minimal panel size in order to recover the minimal rate for a known X.

In Section 2 we introduce our basic assumptions and recall the minimax theory derived in Cardot and Johannes [2010] for estimating β non-parametrically given an iid. sample of (Y, X). Assuming an iid. sample of size n of $(Y, Z_1, \dots, Z_L)$ we derive in Section 3 a lower bound in terms of both, n and L, for a maximal weighted risk. We propose an estimator based on dimension reduction and thresholding techniques that can attain the lower bound up to a constant. All proofs can be found in Bereswill and Johannes [2010].

2. Background to the methodology

For sake of simplicity we assume that the measurement errors ε and $\{\xi_{j,\ell}\}_{j\in\mathbb{N}, 1\leqslant \ell\leqslant L}$ are independent and standard normally distributed, i.e, $\Xi_1, \dots, \Xi_L$ are independent Gaussian white noises in $\mathbb{H}$. Furthermore, we suppose that the regressor X is Gaussian with mean zero and a finite second moment, i.e., $\mathbb{E}\|X\|^2 < \infty$, as well as independent of all measurement errors. Taking the expectation after multiplying both sides in (1a) by X we obtain $g := \mathbb{E}[YX] = \mathbb{E}[\langle\beta, X\rangle X] =: \Gamma\beta$, where g belongs to $\mathbb{H}$ and Γ denotes the covariance operator associated with the random function X. In what follows we always assume that there exists in $\mathbb{H}$ a unique solution of the equation $g = \Gamma\beta$, i.e., that g belongs to the range of the strictly positive Γ (c.f. Cardot et al. [2003]). It is well-known that the obtainable accuracy of any estimator of β can essentially be determined by the regularity conditions imposed on both, the slope parameter β and the covariance operator Γ. We formalize now these conditions, which are characterized in this paper by different weighted norms in $\mathbb{H}$ with respect to the pre-specified basis $\{\psi_j\}_{j\geqslant}$.

Given a positive sequence of weights $w := (w_j)_{j\geqslant 1}$ we define the weighted norm $\|f\|_w^2 := \sum_{j\geqslant 1} w_j |\langle f, \psi_j\rangle|^2$, $f \in \mathbb{H}$, the completion $\mathcal{F}_w$ of $\mathbb{H}$ with respect to $\|\cdot\|_w$ and the ellipsoid $\mathcal{F}_w^c := \big\{f \in \mathcal{F}_w : \|f\|_w^2 \leqslant c\big\}$ with radius $c > 0$. Here and subsequently, given strictly positive sequences of weights $\gamma := (\gamma_j)_{j\geqslant 1}$ and $\omega := (\omega_j)_{j\geqslant 1}$ we shall measure the performance of any estimator $\widehat{\beta}$ by its maximal $\mathcal{F}_\omega$-risk over the ellipsoid $\mathcal{F}_\gamma^\rho$ with radius $\rho > 0$, that is $\sup_{\beta\in\mathcal{F}_\gamma^\rho} \mathbb{E}\|\widehat{\beta} - \beta\|_\omega^2$. This general framework allows us with appropriate choices of the basis $\{\psi_j\}_{j\geqslant}$ and the weight sequence ω to cover the estimation not only of the slope function itself (c.f. Hall and Horowitz [2007]) but also of its derivatives as well as the optimal estimation with respect to the mean squared prediction error (c.f. Crambes et al. [2009]). For a more detailed discussion, we refer to Cardot and Johannes [2010]. Furthermore, as usual in the context of ill-posed inverse problems, we link the mapping properties of the covariance operator Γ and the regularity conditions on β. Denote by $\mathcal{N}$ the set of all strictly positive nuclear operators defined on $\mathbb{H}$. Given a strictly positive sequence of weights $\lambda := (\lambda_j)_{j\geqslant 1}$ and a constant $d \geqslant 1$ define the subset $\mathcal{N}_\lambda^d := \{\Gamma \in \mathcal{N} : \|f\|_\lambda^2/d^2 \leqslant \|\Gamma f\|^2 \leqslant d^2\,\|f\|_\lambda^2,\ \forall f \in \mathbb{H}\}$ of $\mathcal{N}$. Notice that $\langle\Gamma\psi_j, \psi_j\rangle \geqslant d^{-1}\lambda_j^{1/2}$ for all $\Gamma \in \mathcal{N}_\lambda^d$, and hence the sequence $(\lambda_j^{1/2})_{j\geqslant 1}$ is necessarily summable. All the results in this paper are derived with respect to the three sequences ω, γ and λ. We do not specify these sequences, but impose from now on the following minimal regularity conditions.

Assumption (A.1). *Let $\omega := (\omega_j)_{j\geqslant 1}$, $\gamma := (\gamma_j)_{j\geqslant 1}$ and $\lambda := (\lambda_j)_{j\geqslant 1}$ be strictly positive sequences of weights with $\gamma_1 = 1$, $\omega_1 = 1$ and $\lambda_1 = 1$ such that γ and $(\gamma_j/\omega_j)_{j\geqslant 1}$ are non decreasing, λ and $(\lambda_j/\omega_j)_{j\geqslant 1}$ are non increasing with $\Lambda := \sum_{j=1}^\infty \lambda_j^{1/2} < \infty$.*

Given a sample size $n \geqslant 1$ and sequences ω, γ and λ satisfying Assumption A.1 define

$$m_n^* := m_n^*(\gamma, \omega, \lambda) := \mathop{\arg\min}_{m\geqslant 1} \Big\{\max\Big(\tfrac{\omega_m}{\gamma_m}, \textstyle\sum_{j=1}^m \tfrac{\omega_j}{n\sqrt{\lambda_j}}\Big)\Big\} \text{ and }$$
$$\delta_n^* := \delta_n^*(\gamma, \omega, \lambda) := \max\Big(\tfrac{\omega_{m_n^*}}{\gamma_{m_n^*}}, \textstyle\sum_{j=1}^{m_n^*} \tfrac{\omega_j}{n\sqrt{\lambda_j}}\Big). \quad (3)$$

If in addition $\triangle := \inf_{n\geqslant 1}\{(\delta_n^*)^{-1}\min(\omega_{m_n^*}\gamma_{m_n^*}^{-1}, \sum_{j=1}^{m_n^*} \omega_j(n\sqrt{\lambda_j})^{-1})\} > 0$, then there exists $C > 0$ depending on $\sigma^2, \rho, d, \triangle$ only such that (c.f. Cardot and Johannes [2010]),

$$\inf_{\breve{\beta}} \inf_{\Gamma\in\mathcal{N}_\lambda^d} \sup_{\beta\in\mathcal{F}_\gamma^\rho} \Big\{\mathbb{E}\|\breve{\beta} - \beta\|_\omega^2\Big\} \geqslant C\,\delta_n^* \quad \text{for all } n \geqslant 1.$$

Assuming an iid. sample $\{(Y^{(i)}, X^{(i)})\}$ of size n of (Y, X), it is natural to consider the estimators $\widetilde{g} := \frac{1}{n}\sum_{i=1}^n Y^{(i)}\, X^{(i)}$ and $\widetilde{\Gamma} := \frac{1}{n}\sum_{i=1}^n \langle\cdot, X^{(i)}\rangle X^{(i)}$ for g and Γ respectively. Given $m \geqslant 1$, we denote by $[\widetilde{\Gamma}]_{\underline{m}}$ the $m\times m$ matrix with generic elements $[\widetilde{\Gamma}]_{j,\ell} := \langle \widetilde{\Gamma}\psi_\ell, \psi_j\rangle = n^{-1}\sum_{i=1}^n \langle X^{(i)}, \psi_\ell\rangle\langle X^{(i)}, \psi_j\rangle$, and by $[\widetilde{g}]_{\underline{m}}$ the m vector with elements $[\widetilde{g}]_\ell := \langle \widetilde{g}, \psi_\ell\rangle = n^{-1}\sum_{i=1}^n Y^{(i)}\langle X^{(i)}, \psi_\ell\rangle$, $1 \leqslant j, \ell \leqslant m$. Obviously, if $[\widetilde{\Gamma}]_{\underline{m}}$ is non singular then $[\widetilde{\Gamma}]_{\underline{m}}^{-1}[\widetilde{g}]_{\underline{m}}$ is a least squares estimator of the vector $[\beta]_{\underline{m}}$ with elements $\langle \beta, \psi_\ell\rangle, 1 \leqslant \ell \leqslant m$. The estimator of β consists now in thresholding this projection estimator, that is,

$$\widetilde{\beta}_m := \sum_{j=1}^m \widetilde{[\beta]}_j \psi_j \quad \text{with} \quad \widetilde{[\beta]}_{\underline{m}} := \begin{cases} [\widetilde{\Gamma}]_{\underline{m}}^{-1}[\widetilde{g}]_{\underline{m}}, & \text{if } [\widetilde{\Gamma}]_{\underline{m}} \text{ is non-singular} \\ & \text{ and } \|[\widetilde{\Gamma}]_{\underline{m}}^{-1}\| \leqslant n, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Under Assumption A.1 and $\sup_{m\geqslant 1} m^4\lambda_m/\gamma_m < \infty$ it is shown in Cardot and Johannes [2010] that there exists $C > 0$ depending on $\sigma^2, \rho, d, \Lambda$ only such that

$$\sup_{\Gamma\in\mathcal{N}_\lambda^d}\ \sup_{\beta\in\mathcal{F}_\gamma^\rho} \left\{ \mathbb{E}\|\widetilde{\beta}_{m_n^*} - \beta\|_\omega^2 \right\} \leqslant C\, \delta_n^*\ ,$$

where the dimension parameter m_n^* is given in (4).

Examples of rates. We compute in this section the minimal rate δ_n^* for two standard configurations for γ, ω, and λ. In both examples, we take $\omega_j = j^{2s}$, $s \in \mathbb{R}$, for $j \geqslant 1$. Here and subsequently, we write $a_n \lesssim b_n$ if there exists $C > 0$ such that $a_n \leqslant C\, b_n$ for all $n \in \mathbb{N}$ and $a_n \sim b_n$ when $a_n \lesssim b_n$ and $b_n \lesssim a_n$ simultaneously.

(*p-p*) For $j \geqslant 1$ let $\gamma_j = j^{2p}$, $p > 0$, and $\lambda_j = j^{-2a}$, $a > 1$, then Assumption A.1 holds, if $-a < s < p$. It is easily seen that $m_n^* \sim n^{1/(2p+a+1)}$ if $2s + a > -1$, $m_n^* \sim n^{1/[2(p-s)]}$ if $2s + a < -1$ and $m_n^* \sim (n/\log(n))^{1/[2(p-s)]}$ if $a + 2s = -1$. The minimal rate δ_n^* attained by the estimator is $\max(n^{-(2p-2s)/(a+2p+1)}, n^{-1})$, if $2s + a \neq -1$ (and $\log(n)/n$ if $2s + a = -1$). Since an increasing value of a leads to a slower minimal rate, it is called degree of ill-posedness (c.f. Natterer [1984]). Moreover, the case $0 \leqslant s < p$ can be interpreted as the L^2-risk of an estimator of the s-th derivative of β. On the other hand $s = -a/2$ corresponds to the mean-prediction error (c.f. Cardot and Johannes [2010]).

(*p-e*) For $j \geqslant 1$ let $\gamma_j = j^{2p}$, $p > 0$, and $\lambda_j = \exp(-j^{2a})$, $a > 0$, where Assumption A.1 holds, if $p > s$. Then $m_n^* \sim (\log n - \frac{2p+(2a-1)_+}{2a}\log(\log n))^{1/(2a)}$ with $(q)_+ := \max(q, 0)$. Thereby, $(\log n)^{-(p-s)/a}$ is the minimal rate attained by the estimator.

3. The effect of noisy observations of the regressor

In order to formulate the lower bound below let us define for all $n, L \geqslant 1$ and $\varsigma \geqslant 0$

$$m_{n,L,\varsigma}^* := m_{n,L,\varsigma}^*(\gamma, \omega, \lambda) := \mathop{\arg\min}_{m\geqslant 1} \left\{ \max\left(\tfrac{\omega_m}{\gamma_m}, \textstyle\sum_{j=1}^m \tfrac{\omega_j}{n\sqrt{\lambda_j}}, \textstyle\sum_{j=1}^m \tfrac{\varsigma^2\,\omega_j}{Ln\lambda_j} \right) \right\} \text{ and}$$

$$\delta_{n,L,\varsigma}^* := \delta_{n,L,\varsigma}^*(\gamma, \omega, \lambda) := \max\left(\frac{\omega_{m_{n,L,\varsigma}^*}}{\gamma_{m_{n,L,\varsigma}^*}}, \sum_{j=1}^{m_{n,L,\varsigma}^*} \frac{\omega_j}{n\sqrt{\lambda_j}}, \sum_{j=1}^{m_{n,L,\varsigma}^*} \frac{\varsigma^2\,\omega_j}{Ln\lambda_j} \right). \tag{5}$$

The lower bound given below needs the following assumption.

Assumption (A.2). *Let ω, γ and λ be sequences such that*

$$0 < \triangle := \inf_{L,n\geqslant 1}\left\{(\delta^*_{n,L,\varsigma})^{-1}\min\left(\frac{\omega_{m^*_{n,L,\varsigma}}}{\gamma_{m^*_{n,L,\varsigma}}}, \sum_{j=1}^{m^*_{n,L,\varsigma}}\frac{\omega_j}{n\sqrt{\lambda_j}}, \sum_{j=1}^{m^*_{n,L,\varsigma}}\frac{\varsigma^2\,\omega_j}{Ln\lambda_j}\right)\right\}\leqslant 1.$$

Theorem (Lower bound). *If the sequences ω, γ and λ satisfy Assumptions A.1 - A.2, then there exists $C > 0$ depending on $\sigma^2, \varsigma^2, \rho, d,$ and $\triangle$ only such that*

$$\inf_{\breve\beta}\inf_{\Gamma\in\mathcal{N}^d_\lambda}\sup_{\beta\in\mathcal{F}^\rho_\gamma}\left\{\mathbb{E}\|\breve\beta-\beta\|^2_\omega\right\}\geqslant C\,\delta^*_{n,L,\varsigma}\qquad \textit{for all } n, L\geqslant 1.$$

Observe that the lower rate $\delta^*_{n,L,\varsigma}$ is never faster than the lower rate δ^*_n for known X defined in (3). Clearly, we recover δ^*_n for all $L \geqslant 1$ in case $\varsigma = 0$. On the other hand given an iid. sample $\{(Y^{(i)}, Z^{(i)}_1, \dots, Z^{(i)}_L)\}$ of size n of $(Y, Z_1, \dots, Z_L)$ we define estimators for the elements $[g]_j := \langle g, \psi_j\rangle$ and $[\Gamma]_{k,j} := \langle\Gamma\psi_k, \psi_j\rangle$, $k, j \geqslant 1$, respectively as follows

$$\widehat{[g]}_j := \frac{1}{n}\sum_{i=1}^n Y^i\frac{1}{L}\sum_{\ell=1}^L Z^{(i)}_{j,\ell}, \quad\text{and}\quad \widehat{[\Gamma]}_{k,j} := \frac{1}{n}\sum_{i=1}^n\frac{1}{L(L-1)}\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1\neq\ell_2}}^L Z^{(i)}_{j,\ell_1}Z^{(i)}_{k,\ell_2}. \tag{6}$$

We replace in definition (4) then the unknown matrix $[\widetilde\Gamma]_{\underline{m}}$ and vector $[\widetilde g]_{\underline{m}}$ respectively by the matrix $\widehat{[\Gamma]}_{\underline{m}}$ with elements $\widehat{[\Gamma]}_{k,j}$ and the vector $\widehat{[g]}_{\underline{m}}$ with elements $\widehat{[g]}_j$, that is,

$$\widehat\beta_m := \sum_{j=1}^m \widehat{[\beta]}_j\psi_j \quad\text{with}\quad \widehat{[\beta]}_{\underline{m}} := \begin{cases}\widehat{[\Gamma]}_{\underline{m}}^{-1}\widehat{[g]}_{\underline{m}}, & \text{if } \widehat{[\Gamma]}_{\underline{m}} \text{ is non-singular}\\ & \text{ and } \|\widehat{[\Gamma]}_{\underline{m}}^{-1}\|\leqslant n,\\ 0, & \text{otherwise.}\end{cases} \tag{7}$$

The next theorem establishes the minimax-optimality of the estimator $\widehat\beta_m$ provided the dimension parameter m is chosen appropriate, i.e $m := m^*_{n,L,\varsigma}$ given in (5).

Theorem (Upper bound). *If Assumptions A.1 - A.2 and $\sup_{m\geqslant 1} m^4\lambda_m\gamma_m^{-1} < \infty$ are satisfied, then there exists $C > 0$ depending on $\sigma^2, \varsigma^2, \rho, d, \Lambda$ only such that*

$$\sup_{\Gamma\in\mathcal{N}^d_\lambda}\sup_{\beta\in\mathcal{F}^\rho_\gamma}\left\{\mathbb{E}\|\widehat\beta_{m^*_{n,L,\varsigma}}-\beta\|^2_\omega\right\}\leqslant C\,\delta^*_{n,L,\varsigma}\qquad \textit{for all } n\geqslant 1, L\geqslant 2 \textit{ and } \varsigma\geqslant 0.$$

Examples of rates (continued). Suppose first that the panel size $L \geqslant 2$ is constant and $\varsigma > 0$. In example (*p-p*) if $2s+2a+1 > 0$ it is easily seen that $m^*_{n,L,\varsigma} \sim n^{1/(2p+2a+1)}$ and the minimal rate attained by the estimator is $\delta^*_{n,L,\varsigma} \sim n^{-(2p-2s)/(2a+2p+1)}$. Let us compare this rate with the minimal rates in case of a functional linear model (FLM) with known regressor and in case of an indirect regression model (IRM) given by the covariance operator Γ and Gaussian white noise $\dot W$, i.e., $g_n = \Gamma\beta + n^{-1/2}\dot W$ (c.f. Hoffmann and Reiß [2008]). The minimal rate in the FLM with known X is $n^{-2(p-s)/(a+2p+1)}$, while $n^{-2(p-s)/(2a+2p+1)}$ is the minimal rate in the IRM. We see that in a FLM with known X the covariance operator Γ has the *degree of ill-posedness* a while it has in a FLM with noisy observations of X and in the IRM a *degree of ill-posedness* $2a$. In other words only in a FLM with known regressor we do not face the complexity of an inversion of Γ but only of its square root $\Gamma^{1/2}$. The same remark holds true in the example (*p-e*), but the minimal rate is the same in all three cases due to the fact that for $\lambda_j \sim \exp(-r|j|^{2a})$ the dependence of the minimal rate on the value r is hidden in the constant. However, it is rather surprising that in this situation a panel of size $L = 2$ is sufficient to recover the minimal but logarithmic rate when X is known. In contrast, in example (*p-p*) the minimal rate for known X can only be attained in the presence of noise in the regressor if the panel size satisfies $L_n^{-1} = O(n^{-a/(a+2p+1)})$ as the sample size n increases, since $\delta^*_{n,L,\varsigma} \sim \max(n^{-(2p-2s)/(a+2p+1)}, (L_n n)^{-(2p-2s)/(2a+2p+1)})$.

References

1. M. Bereswill and J. Johannes. On the effect of noisy observations of the regressor in a functional linear model. Technical report, Université catholique de Louvain, 2010.
2. D. Bosq. *Linear Processes in Function Spaces.*, volume 149 of *Lecture Notes in Statistics.* Springer-Verlag, 2000.
3. H. Cardot and J. Johannes. Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, 101(2):395–408, 2010.
4. H. Cardot, F. Ferraty, and P. Sarda. Functional linear model. *Statistics & Probability Letters*, 45:11–22, 1999.
5. H. Cardot, F. Ferraty, and P. Sarda. Spline estimators for the functional linear model. *Statistica Sinica*, 13:571–591, 2003.
6. H. Cardot, A. Mas, and P. Sarda. Clt in functional linear regression models. *Probability Theory and Related Fields*, 138:325–361, 2007.
7. C. Crambes, A. Kneip, and P. Sarda. Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, 37(1):35–72, 2009.
8. S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.
9. F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations.* Springer-Verlag, London, 2006.
10. P. Hall and J. L. Horowitz. Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91, 2007.
11. M. Hoffmann and M. Reiß. Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, 36(1):310–336, 2008.
12. B. D. Marx and P. H. Eilers. Generalized linear regression on sampled signals and curves: a p-spline approach. *Technometrics*, 41:1–13, 1999.
13. H.-G. Müller and U. Stadtmüller. Generalized functional linear models. *The Annals of Statistics*, 33(2):774–805, 2005.
14. F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Analysis*, 18:29–37, 1984.
15. J. Ramsay and B. Silverman. *Functional Data Analysis.* Springer, New York, second ed. edition, 2005.
16. F. Yao, H.-G. Müller, and J.-L. Wang. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903, 2005.