Journal of Multivariate Analysis 101 (2010) 395–408

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Thresholding projection estimators in functional linear models

Hervé Cardot [a,*], Jan Johannes [b]

[a] Université de Bourgogne, Institut de Mathématiques de Bourgogne, 9 Av. Alain Savary, 21078 Dijon Cedex, France

[b] Universität Heidelberg, Institut für Angewandte Mathematik, Im Neuenheimer Feld, 294, D-69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Available online 24 March 2009

AMS 2000 subject classifications:
primary 62J05
secondary 62G20
62G08

Keywords:
Derivatives estimation
Galerkin method
Linear inverse problem
Mean squared error of prediction
Optimal rate of convergence
Hilbert scale
Sobolev space

ABSTRACT

We consider the problem of estimating the regression function in functional linear regression models by proposing a new type of projection estimators which combine dimension reduction and thresholding. The introduction of a threshold rule allows us to get consistency under broad assumptions as well as minimax rates of convergence under additional regularity hypotheses. We also consider the particular case of Sobolev spaces generated by the trigonometric basis which permits us to get easily mean squared error of prediction as well as estimators of the derivatives of the regression function. We prove that these estimators are minimax and rates of convergence are given for some particular cases.

1. Introduction

Functional data analysis [1,2] is a topic of growing interest in statistics and many applications in chemometrics [3], finance [4], biometry or climatology [5] are now dealing with the functional linear model. This model is useful to estimate or predict a scalar random variable, say $Y \in \mathbb{R}$, thanks to a random function denoted by X. We assume in the following that Y and X are centered random variables and, without loss of generality, that the random function X takes values in $L^2[0,1]$, the space of square integrable functions defined on $[0,1]$ endowed with its usual inner product $\langle f,g\rangle = \int_0^1 f(t)g(t)\mathrm{d}t$ and associated norm $\|f\| = \langle f,f\rangle^{1/2}, f,g \in L^2[0,1]$. The functional linear model is then defined by

$$Y = \int_0^1 \beta(t)X(t)\mathrm{d}t + \sigma\epsilon, \quad \sigma > 0, \tag{1.1}$$

where the function $\beta(t)$ is called the regression or slope function and the error term ϵ is supposed to be centered $\mathbb{E}(\epsilon) = 0$ and not correlated with X: $\forall\, t \in [0,1],\ \mathbb{E}(X(t)\epsilon) = 0$.

Assuming that X has a finite second moment, i.e. $\mathbb{E}\|X\|^2 = \int_0^1 \mathbb{E}|X(t)|^2\mathrm{d}t < \infty$, one can define the covariance operator of X, say Γ. This operator is defined on $L^2[0,1]$ as follows: for any function $f \in L^2[0,1]$,

$$\Gamma f(s) = \int_0^1 \mathrm{cov}(X(t),X(s))f(t)\,\mathrm{d}t, \quad \forall s \in [0,1]. \tag{1.2}$$

* Corresponding author.
E-mail addresses: herve.cardot@u-bourgogne.fr (H. Cardot), johannes@statlab.uni-heidelberg.de (J. Johannes).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.03.001

It is well known (see *e.g.* [6]) that the regression function β satisfies the following moment equation

$$g(s) := \mathbb{E}[YX(s)] = [\Gamma\beta](s), \quad s \in [0, 1], \tag{1.3}$$

where g belongs to $L^2[0, 1]$. Since Γ is a nonnegative nuclear operator [7] a continuous generalized inverse of Γ does not exist as long as the range of the operator Γ is an infinite dimensional subspace of $L^2[0, 1]$. Consequently inverting Eq. (1.3) to recover β can be seen as an ill-posed inverse problem. Cardot et al. [8] provide a necessary and sufficient condition for the existence of a unique solution of Eq. (1.3).

Assumption 1.1. The covariance operator Γ of the random function X is injective and the function $g = \mathbb{E}[YX]$ belongs to the range $\mathcal{R}(\Gamma)$ of Γ.

Under this assumption, the covariance operator Γ admits a discrete spectral decomposition given by a sequence $(\lambda_j)_{j\in\mathbb{N}}$ of strictly positive eigenvalues and a sequence of corresponding orthonormal eigenfunctions $\{\phi_j\}_{j\in\mathbb{N}}$. Then, the normal equation (1.3) can be rewritten as follows

$$\beta = \sum_{j\in\mathbb{N}} \frac{g_j}{\lambda_j} \cdot \phi_j \quad \text{with } g_j := \langle g, \phi_j\rangle,\ j \in \mathbb{N}. \tag{1.4}$$

It is well known that, even in the case of a priori known eigenvalues $\{\lambda_j\}$ and eigenfunctions $\{\phi_j\}$, replacing in (1.4) the unknown function g by a consistent estimator $\widehat{g}$ does in general not lead to a consistent estimator of β. To be more precise, since the sequence $(\lambda_j)_{j\in\mathbb{N}}$ tends to zero, $\mathbb{E}\|\widehat{g} - g\|^2 = o(1)$ does generally not imply $\sum_{j\in\mathbb{N}} |\lambda_j|^{-2} \cdot \mathbb{E}|\langle\widehat{g} - g, \phi_j\rangle|^2 = o(1)$. Consequently, the estimation in functional linear model is called ill-posed and additional regularity assumptions on the regression function β are necessary in order to obtain a uniform rate of convergence (cf. [9]).

The objective is to estimate the regression function β, as well as its derivatives, when observing a sample (Y_i, X_i) of n i.i.d. realizations of (Y, X). We can define the empirical estimators of g and Γ respectively as follows

$$\widehat{g} := \frac{1}{n}\sum_{i=1}^{n} Y_i X_i \quad \text{and} \quad \widehat{\Gamma} := \frac{1}{n}\sum_{i=1}^{n}\langle X_i, \cdot\rangle X_i. \tag{1.5}$$

The main class of estimation procedures studied in the statistical literature are based on principal components regression and consist in reducing the dimension by inverting Eq. (1.3) in the finite dimension space generated by the eigenfunctions of $\widehat{\Gamma}$ associated to the largest eigenvalues (see *e.g.* [10,3,6,11] or [12] in the context of generalized linear models).

The second important class of estimators relies on minimizing a penalized least squares criterion which can be seen as a generalization of the ridge regression. Marx and Eilers [13], and Cardot et al. [8] proposed B-splines expansion of the regression function with a penalty dealing with the squared norm of a fixed order derivative of the estimators. More recently Crambes et al. [14] proposed a spline smoothing decomposition with the same type of penalty and proved the optimality of their estimators according to a criterion that can be interpreted as a squared error of prediction. Note that this question has given rise recently to numerous publications in the machine learning community with similar ideas based on reproducing kernel Hilbert spaces (RKHS) and Tikhonov regularization (see e.g. [15,16] and the references therein).

Borrowing ideas from the inverse problems community [17,18] we propose in this article a new class of estimators which rely on dimension reduction by projecting the data onto some basis of orthonormal functions and threshold techniques that allow us to control the accuracy of the estimator. More precisely, let us consider a set of orthonormal functions such as wavelet or trigonometric basis denoted by $\{\psi_1, \ldots, \psi_m, \ldots\}$ which forms a basis of $L^2[0, 1]$. Given a dimension $m \geq 1$, we denote by $[\widehat{\Gamma}]_{\underline{m}}$ the $m \times m$ matrix with generic elements $\langle\widehat{\Gamma}\psi_\ell, \psi_j\rangle, j, \ell = 1, \ldots, m$ and by $[\widehat{g}]_{\underline{m}}$ the m vector with elements $\langle\widehat{g}, \psi_\ell\rangle$, $\ell = 1, \ldots, m$. We can first remark, that the least squares estimator of β obtained with the projections of the X_i onto Ψ_m, the subspace of $L^2[0, 1]$ spanned by the functions $\{\psi_1, \ldots, \psi_m\}$, is simply given, when $[\widehat{\Gamma}]_{\underline{m}}$ is nonsingular, by $([\widehat{\Gamma}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}})^{t}[\psi]_{\underline{m}}(\cdot)$ where $[\psi]_{\underline{m}}(\cdot) = (\psi_1(\cdot), \ldots, \psi_m(\cdot))^{t}$. Our estimator, in its simplest form, consists in thresholding this projection estimator when, roughly speaking, the norm of the inverse of the matrix $[\widehat{\Gamma}]_{\underline{m}}$ is too large. More precisely, introducing a threshold value γ which will depend on m and n we propose to estimate β as follows

$$\widehat{\beta}(t) = \sum_{\ell=1}^{m} \widehat{\beta}_\ell \cdot \mathbb{1}\{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\| \leq \gamma\} \cdot \psi_\ell(t), \quad t \in [0, 1], \tag{1.6}$$

where the $\widehat{\beta}_\ell$ are the generic elements of the vector of coordinates obtained by least squares projection and $\mathbb{1}$ is the indicator function. This new thresholding step can be seen as an improvement of the estimator proposed by Ramsay and Dalzell [19] which was built by projecting the data onto finite dimensional basis of functions. From an inverse problem perspective this approach is similar to the linear Galerkin procedure [20] or [9] defined as follows, $\beta^m \in \Psi_m$ denotes a Galerkin solution of the operator equation $g = \Gamma\beta$ when

$$\|g - \Gamma\beta^m\| \leqslant \|g - \Gamma\tilde{\beta}\|, \quad \forall\tilde{\beta} \in \Psi_m. \tag{1.7}$$