Penalized spline approaches for functional logit regression

- M. Carmen Aguilera-Morillo; Ana M. Aguilera; Manuel Escabias; Mariano J. Valderrama

- Penalized spline approaches for functional logit regression

ORIGINAL PAPER

Penalized spline approaches for functional logit regression

**M. Carmen Aguilera-Morillo · Ana M. Aguilera ·
Manuel Escabias · Mariano J. Valderrama**

Abstract The problem of multicollinearity associated with the estimation of a functional logit model can be solved by using as predictor variables a set of functional principal components. The functional parameter estimated by functional principal component logit regression is often nonsmooth and then difficult to interpret. To solve this problem, different penalized spline estimations of the functional logit model are proposed in this paper. All of them are based on smoothed functional PCA and/or a discrete penalty in the log-likelihood criterion in terms of B-spline expansions of the sample curves and the functional parameter. The ability of these smoothing approaches to provide an accurate estimation of the functional parameter and their classification performance with respect to unpenalized functional PCA and LDA-PLS are evaluated via simulation and application to real data. Leave-one-out cross-validation and generalized cross-validation are adapted to select the smoothing parameter and the number of principal components or basis functions associated with the considered approaches.

M.C. Aguilera-Morillo (✉) · M. Escabias · M.J. Valderrama
Departamento de Estadística e Investigación Operativa, Facultad de Farmacia, Campus Universitario de Cartuja, 18071 Granada, Spain
e-mail: caguilera@ugr.es

A.M. Aguilera
Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Campus de Fuentenueva, 18071 Granada, Spain

1 Introduction

A part of the literature has recently been concerned with functional data in a wide variety of statistical problems, and with developing procedures based on smoothing techniques. A functional data set provides information about functions (curves, surfaces, etc.) varying over a continuum. The argument of the sample functions is often time, but may also be a different magnitude as spatial location, wavelength or probability. A magistral compilation of models working with sample curves and interesting applications in different fields are collected in Ramsay and Silverman (2005) and Ramsay and Silverman (2002), respectively.

The aim of the functional logit model (FLM) is to predict a binary response variable from a functional predictor and also to interpret the relationship between the response and the predictor variables. In the last years, the FLM was applied in different contexts. A FLM was applied to predict if human foetal heart rate responds to repeated vibroacoustic stimulation (Ratcliffe et al. 2002). The FLM was considered in the more general framework of functional generalized linear models in James (2002). A nonparametric estimation procedure of the generalized functional linear model for the case of sparse longitudinal predictors was proposed in Müller (2005). This extension included functional binary regression models for longitudinal data and was illustrated with data on primary biliary cirrhosis. An alternative nonparametric classification method was studied in Ferraty and Vieu (2003).

In order to reduce the infinite dimension of the functional predictor and to solve the multicollinearity problem associated with the estimation of the FLM, a reduced number of functional principal components can be used as predictor variables to provide accurate estimation of the functional parameter (Escabias et al. 2004). A climatological application to establish the relationship between the risk of drought and time evolution of temperatures was carried out by Escabias et al. (2005). The relationship between lupus flares and stress level was analyzed by using a principal component logit model in Aguilera et al. (2008). A functional PLS based solution was also proposed by Escabias et al. (2007). The problem associated with these approaches is that in many cases the estimated functional parameter is not smooth and therefore difficult to interpret. The main objective of this paper is to solve this problem by introducing different penalties based on P-splines.

The functional linear model was the first regression model extended to the case of functional data. In order to estimate an accurate functional parameter, a smoothing estimation approach based on penalizing the least squares criterion in terms of the squared norm of a B-spline expansion of the functional parameter was introduced by Cardot et al. (2003). A smoothed principal component regression based on ordinary least squares regression on the projection of the covariables on a set of eigenfunctions was also considered. When the functional predictor is corrupted by some error, the functional parameter was estimated by total least squares by using smoothing splines (continuous spline penalty based on the integral of the squared second derivative of the functional parameter) (Cardot et al. 2007). Two versions of functional PCR for scalar response using B-splines and discrete roughness penalty were proposed in Reiss and Ogden (2007). In one of them, the penalty is introduced in the construction of the principal components. In the other one, a penalized likelihood estimation

is considered. The smoothing parameter was found by fitting a linear mixed model. These penalized PCR approaches did not consider the functional form of the sample paths but only the approximation in terms of basis functions of the functional parameter. When both the response and the predictor variables are functional, the idea of discrete roughness penalties based on the absolute values of the basis function coefficient differences (corresponding to the LASSO) and the squares of these differences (according to the P-spline methodology) was extended to the functional linear model setting by penalizing the interpretable directions of the regression surface in Harezlak et al. (2007). From a Bayesian point of view, approaches to control the modes of variation in a set of noisy and sparse curves were proposed by van der Linde (2008) where Demmler–Reinsch basis was used to get smooth weight functions in the functional PCA estimation.

In the general context of functional generalized linear models (FGLM), different penalized likelihood estimations with B-spline basis were proposed to solve the roughness problem of the functional parameter. The FGLM with P-spline penalty in the log-likelihood criterion was developed in Marx and Eilers (1999). The benefits of this functional model were compared with functional PLS and PCR. A penalized estimation of the functional parameter via penalized log-likelihood was proposed by Cardot and Sarda (2005). This estimation is quite similar to the one provided by Marx and Eilers (1999) with the main difference coming from the continuous penalty that was expressed as the norm of the derivative of given order of the function. A practical mechanism to combine the GLM via penalized log-likelihood, the general additive models (Hastie and Tibshirani 1990) and the varying-coefficient model (Hastie and Tibshirani 1993) into a general additive structure was introduced by Eilers and Marx (2002).

In this work, we propose four different methods based on penalized spline (P-spline) estimation of the functional logit regression model by considering the functional form of the sample paths and the functional parameter in terms of B-spline basis expansions. The considered approaches are based on smoothed functional principal component logit regression (FPCLOR) and functional logit regression via penalized log-likelihood.

In the FPCLOR context, three different versions of penalized estimation approaches based on smoothed functional principal component analysis (FPCA) are introduced. On the one hand, FPCA of P-spline approximation of sample curves (Method II) is performed. On the other hand, a discrete P-spline penalty that penalizes the roughness of the principal component weight functions is included in the own formulation of FPCA (Method III). The third smoothed FPCLOR approach is carried out by introducing the penalty in the likelihood estimation of the functional parameter in terms of a reduced set of functional principal components (Method IV). Moreover, direct P-spline likelihood estimation in terms of B-spline functions is also considered (Method V).

The good performance of the proposed methods with respect to non-penalized FPCLOR (Method I) and LDA-PLS is evaluated via two different data simulations, a functional version of the well-known waveform data and a smooth principal component reconstruction of the Ornstein–Uhlenbeck process. This study is completed with an application to real data whose aim is to estimate the quality of cookies (good

or bad) on the basis of the curves of resistance of dough during the kneading process (functional data classification).

2 Functional logit model

The main objective of this paper consists in estimating the link between a binary random variable Y and a functional predictor $X = \{X(t)\}_{t \in T}$. It will be assumed without loss of generality that X is a centered second order stochastic process whose sample paths belong to the space $L_2(T)$ of square integrable functions with the usual inner product defined by $\langle f, g \rangle = \int_T f(t)g(t) dt$. This means that $E(X(t)) = 0, \forall t \in T$.

Let $\{x_1(t), x_2(t), \ldots, x_n(t)\}$ be a sample of the functional variable X and $\{y_1, y_2, \ldots, y_n\}$ be a random sample of Y associated with them. That is, $y_i \in \{0, 1\}$, $i = 1, \ldots, n$. The functional logistic regression model is given by

$$y_i = \pi_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

where π_i is the expectation of Y given $x_i(t)$ modeled as

$$\pi_i = P\big[Y = 1 \big| \{x_i(t) : t \in T\}\big] = \frac{\exp\{\alpha + \int_T x_i(t)\beta(t) dt\}}{1 + \exp\{\alpha + \int_T x_i(t)\beta(t) dt\}}, \quad i = 1, \ldots, n, \quad (1)$$

with α being a real parameter, $\beta(t)$ a functional parameter, and $\{\varepsilon_i : i = 1, \ldots, n\}$ independent errors with zero mean. The logit transformations can be expressed as

$$l_i = \ln\left[\frac{\pi_i}{1 - \pi_i}\right] = \alpha + \int_T x_i(t)\beta(t) dt, \quad i = 1, \ldots, n. \quad (2)$$

In the functional logit model, we have to take into account different aspects. Firstly, we cannot continuously observe the functional form of the sample paths. As much we can observe each sample curve $x_i(t)$ in a finite set of discrete sampling points $\{t_{i0}, t_{i1}, \ldots, t_{im_i} \in T, \ i = 1, \ldots, n\}$, so that the sample information is given by the vectors $x_i = (x_{i0}, \ldots, x_{im_i})'$, with x_{ik} being the observed value for the ith sample path $x_i(t)$ at time t_{ik} ($k = 0, \ldots, m_i$). Secondly, it is impossible to estimate the infinite functional parameter with a finite number of observations n. In order to solve at the same time the two questions, a functional estimation approach based on approximating the sample paths and the functional parameter in terms of basis functions was proposed (Escabias et al. 2007). Different basis such as trigonometric functions (see Aguilera et al. 1995 and Ratcliffe et al. 2002), cubic spline functions (see Aguilera et al. 1996 and Escabias et al. 2005), or wavelet functions (see Ocaña et al. 2008) can be used depending on the nature of the functional predictor sample paths.

Let us consider that both the sample curves and the functional parameter are approximated as a weighted sum of basis functions as follows:

$$x_i(t) = \sum_{j=1}^{p} a_{ij}\phi_j(t), \qquad \beta(t) = \sum_{k=1}^{p} \beta_k \phi_k(t), \quad (3)$$

with p being the number of basis functions. Choosing the order of the expansion p is an important problem. If p is increased, the fit to the data is better, but we risk fitting

noise or variation that affects the raw data. On the other hand, if p is too small, we may miss some important characteristics of the underlying smooth function.

Then, the FLM (2) turns into a multiple logit model whose design matrix is the product between the matrix of basis coefficients of the sample paths and the matrix of inner products between basis functions (Escabias et al. 2004). So, the logit transformations in matrix form are given by

$$L = X\beta, \tag{4}$$

where $L = (l_1, \ldots, l_n)$ is the vector of logit transformations, $X = (\mathbf{1} | A\Psi)$, with $A = (a_{ij})_{n \times p}$ being the matrix of basis coefficients of the sample paths, $\Psi = (\psi_{jk})_{p \times p}$ the matrix of inner products between basis functions ($\psi_{jk} = \int_T \phi_j(t)\phi_k(t)\, dt$), $\mathbf{1} = (1, \ldots, 1)'$ an n-dimensional vector of ones, and $\beta = (\beta_1, \ldots, \beta_p)'$ the vector of basis coefficients of $\beta(t)$.

In order to estimate the multiple logit model (4), we must first approximate the basis coefficients of each sample curve from its discrete time observations (rows of matrix A). When the sample curves are smooth and observed with error, least squares approximation in terms of B-spline basis is an appropriate solution for the problem of reconstructing their true functional form. Other alternatives to B-spline approximation are techniques such as interpolation or projection in a finite-dimensional space generated by basis functions. More recently, nonparametric techniques were used for approximating functional data (Ferraty and Vieu 2006).

B-splines are constructed from polynomial pieces joined at a set of knots. Once the knots are given, B-splines can be evaluated recursively for any degree of the polynomial by using a numerically stable algorithm (De Boor 2001). Considering the least squares approximation in terms of B-spline basis, the vector of basis coefficients of each sample curve that minimizes the least squares error $(x_i - \Phi_i a_i)'(x_i - \Phi_i a_i)$ is given by $\hat{a}_i = (\Phi_i'\Phi_i)^{-1}\Phi_i'x_i$, with $\Phi_i = (\phi_j(t_{ik}))_{m_i \times p}$ and $a_i = (a_{i1}, \ldots, a_{ip})'$. These approximated sample curves are known as regression splines. The choice of the number of knots is an important problem when working with regression splines because they do not control the degree of smoothness of the estimated curve. If too many knots are selected, you have an overfitting of the data. On the other hand, too few knots provide an underfitting. This problem is solved in this paper by using penalized splines. In this case, the smoothness of the approximated curve is controlled by the smoothing parameter.

2.1 Penalized estimation with basis expansions

The log-likelihood function for the multiple model (4) is given by

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \ln(1 - \pi_i) + \sum_{i=1}^{n} y_i \ln\left(\frac{\pi_i}{1 - \pi_i}\right)$$

$$= -\sum_{i=1}^{n} \ln\left(1 + \exp\left(\sum_{j=0}^{p} X_{ij}\beta_j\right)\right) + \sum_{j=0}^{p}\left(\sum_{i=1}^{n} y_i X_{ij}\right)\beta_j. \tag{5}$$

Then, the likelihood equations in matrix form are

$$y'X = \hat{\pi}'X,$$

where $y = (y_1, \ldots, y_n)'$, $\hat{\pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_n)'$ is the vector of likelihood estimators of $\pi = (\pi_1, \ldots, \pi_n)'$, with $\hat{\pi}_i = \frac{\exp(\sum_{j=0}^{p} X_{ij}\hat{\beta}_j)}{1+\exp(\sum_{j=0}^{p} X_{ij}\hat{\beta}_j)}$ and $\hat{\beta}_j$ the likelihood estimators of the basis coefficients of the functional parameter $\beta(t)$ in the FLM. Solving the likelihood equations by mean of the iterative Newton–Raphson method, the vector of basis coefficients of the functional parameter at iteration t is given by

$$\beta^{(t)} = \beta^{(t-1)} + \left[X' \operatorname{Diag}\left(\pi_i^{(t-1)}\left(1 - \pi_i^{(t-1)}\right)\right) X \right]^{-1} X'\left(y - \pi_i^{(t-1)}\right). \qquad (6)$$

The maximum likelihood estimate of the parameters of the logit model can be calculated by iterative reweighted least squares as the limit of a sequence of weighted least squares estimates, where the weight matrix changes each cycle. See Agresti (1990) for a detailed study of this least squares procedure.

The estimation of this model is affected by multicollinearity due to the high correlation between the columns of the design matrix. On the one hand, this problem can be solved by logit regression of the response on a set of uncorrelated variables as, for example, principal components. On the other hand, the problem can be solved by using a penalized estimation of the regression coefficients based on the differences of order d between adjacent coefficients (Le Cessie and Van Houwelingen 1992). In order to obtain a more accurate and smoother estimation of the functional parameter, this methodology is extended in this section to the functional logit model by introducing a penalty in the log-likelihood estimation of the multiple logit model given by Eq. (4). This penalty is based on B-spline basis expansions of the sample curves and the functional parameter, and a simple discrete penalty that measures the roughness of the parameter function by summing the squared dth order differences between adjacent B-spline coefficients (P-spline penalty).

Let us consider the basis expansion of the functional parameter given by Eq. (3). Then, the penalized log-likelihood of the FLM with logit transformation given by (4) is given by

$$\mathcal{L}^*(\lambda, \beta) = \mathcal{L}(\beta) - \frac{\lambda}{2}\beta' P_d \beta,$$

where $\beta = (\beta_1, \ldots, \beta_p)'$ is the vector of basis coefficients of $\beta(t)$, λ is the smoothing parameter, and $P_d = (\Delta^d)'\Delta^d$, with Δ^d the matrix of differences of order d given by the $(p - d) \times p$ matrix

$$\Delta^d = \begin{pmatrix} \binom{d}{0} & \binom{d}{1} & \binom{d}{2} & \cdots & \binom{d}{d} & 0 & 0 & \cdots \\ 0 & \binom{d}{0} & \binom{d}{1} & \cdots & \binom{d}{d-1} & \binom{d}{d} & 0 & \cdots \\ 0 & 0 & \binom{d}{0} & \cdots & \binom{d}{d-2} & \binom{d}{d-1} & \binom{d}{d} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Let us observe that the vector of differences of order d of the vector β is given by $\Delta^d \beta$ and its components are the differences of order 1 of the vector of differences of order $d-1$ given by

$$\sum_{j=0}^{d} \binom{d}{j} \beta_{k+j} \quad k=1,\ldots,p-d.$$

The most common penalty matrix is $P_2 = (\Delta^2)'\Delta^2$, with Δ^2 the $(p-2) \times p$ matrix of differences of order 2 given by

$$\Delta^2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \ldots \\ 0 & 1 & -2 & 1 & 0 & \ldots \\ 0 & 0 & 1 & -2 & 1 & \ldots \\ . & . & . & . & . & \ddots \end{pmatrix}.$$

In this case, the Newton–Raphson solution for the penalized likelihood estimators will be

$$\beta^{(t)} = \beta^{(t-1)} + \left[X' \operatorname{Diag}\left(\pi_i^{(t-1)}\left(1-\pi_i^{(t-1)}\right)\right) X + \lambda P_d \right]^{-1} X'\left(y - \pi_i^{(t-1)}\right). \quad (7)$$

The number of basis functions p and the smoothing parameter λ are selected by means of a double generalized cross validation (double-GCV) procedure (see Sect. 4.4 for more details). Henceforth, this method will be called *Method V*.

3 Penalized estimation of functional principal component logit regression

As said before, the logit regression model given by Eq. (4) is affected by multi-collinearity. In order to solve the problems of high dimension and high correlation between the covariates of this model, a reduction dimension approach based on using as covariates a reduced set of functional principal components of the predictor curves was proposed (Escabias et al. 2004).

In general, the FLM can be rewritten in terms of functional principal components as

$$L = \alpha \mathbf{1} + \Gamma \gamma, \quad (8)$$

where $\Gamma = (\xi_{ij})_{n \times p}$ is a matrix of functional principal components of the sample paths $\{x_1(t), \ldots, x_n(t)\}$, γ is the vector of coefficients of the model and α is the intercept.

An accurate estimation of the functional parameter can be obtained by considering only a set of q optimum principal components as predictor variables, so that $\Gamma = (\xi_{ij})_{n \times q}$ $(q < p)$.

Then, the vector β of basis coefficients is given by $\beta = F\gamma$, where the way of estimating F depends on the kind of functional principal component analysis (FPCA) used to estimate the functional model and the kind of likelihood estimation (penalized or non-penalized). According to it, four different methods are considered in this paper.

3.1 Method I: non-penalized functional principal components logit regression

A simple way to estimate the functional parameter is by means of non-penalized functional logit regression on an optimum set of principal components. This method known as non-penalized functional principal component logit regression (FPCLOR) was performed by Escabias et al. (2004).

In the standard formulation of FPCA, the ith principal component is given by

$$\xi_{ij} = \int_T x_i(t) f_j(t) \, dt, \quad i = 1, \ldots, n, \tag{9}$$

where the weight function or factor loading f_j is obtained by solving

$$\begin{cases} \max_f \ var[\int_T x_i(t) f(t) \, dt] \\ \text{s.t. } \|f\|^2 = 1 \quad \text{and} \quad \int f_\ell(t) f(t) \, dt = 0, \ \ell = 1, \ldots, j - 1. \end{cases}$$

The weight functions f_j are the solutions to the eigenequation $Cf_j = \lambda_j f_j$, with $\lambda_j = var[\xi_j]$ and C being the sample covariance operator defined by $Cf = \int c(\cdot, t) f(t) \, dt$ in terms of the sample covariance function

$$c(s, t) = \frac{1}{n} \sum_{i=1}^{n} x_i(s) x_i(t).$$

In practice, functional PCA has to be estimated from discrete time observations of each sample curve $x_i(t)$ that is approximated in terms of basis functions. If we assume that the sample curves are represented in terms of basis functions as in expression (3), the functional PCA is then equivalent to the multivariate PCA of $A\Psi^{\frac{1}{2}}$ matrix, with $\Psi^{\frac{1}{2}}$ being the square root of the matrix of the inner products between B-spline basis functions (Ocaña et al. 2007). Then, matrix F that provides the relation between the basis coefficients of the functional parameter and the parameters estimated in terms of principal components is given by $F = \Psi_{p \times p}^{-\frac{1}{2}} G_{p \times n}$, where G is the matrix whose columns are the eigenvectors of the sample covariance matrix of $A\Psi^{1/2}$. In this case, the matrix of basis coefficients A is computed by using least squares approximation with B-spline basis and γ is estimated by maximum likelihood without penalty. The optimum number of principal components of the predictor curves used as covariates is chosen by GCV (see Sect. 4.3).

3.2 Method II: FPCLOR on P-spline smoothing of the sample curves

When the sample paths are observed with noise, the estimation of the FLM based on FPCA of regression splines provides a noisy functional parameter. This is because of regression splines do not control the smoothness of the sample paths. In order to quantify the roughness of a curve, a continuous penalty based on the integrated squared second derivative of the function was first introduced by Reinsch (1967). The computation of this continuous penalty in terms of B-splines basis functions was consider in O'Sullivan (1986). This approximation was called smoothing splines. The computational problem of this approach lies in the calculation of the integrals of products of the dth order derivatives between B-spline basis functions. A penalty based on

differences of order d between coefficients of adjacent B-splines (P-spline penalty) was introduced by Eilers and Marx (1996). With this kind of penalty, the choice and position of knots are not determined and it is sufficient to choose a relatively large number of equally spaced basis knots (Ruppert 2002).

Therefore, a penalized estimation of the FLM based on FPCA of the P-spline approximation of the sample curves is proposed. The basis coefficients in terms of B-splines are computed by introducing a discrete penalty in the least squares criterion

$$(x_i - \Phi_i a_i)'(x_i - \Phi_i a_i) + \lambda a_i' P_d a_i,$$

where $P_d = (\Delta^d)' \Delta^d$. The solutions are then given by $\hat{a}_i = (\Phi_i' \Phi_i + \lambda P_d)^{-1} \Phi_i' x_i$. In the P-spline approach, the selection of the smoothing parameter λ is very important because it measures the rate of exchange between fit to the data and variability of the function. The bigger the λ, the smoother the approximated curve. On the other hand, if λ is smaller, the curve tends to become more variable since there is less penalty placed on its roughness. A nonparametric strategy for the choice of the P-spline parameters was performed by Currie and Durban (2002), where mixed model (REML) methods were applied for smoothing parameter selection. In our paper, the smoothing parameter is chosen by leave-one-out cross validation (Sect. 4.1).

Once the P-spline approximation of sample curves has been performed, the multivariate PCA of $A\Psi^{\frac{1}{2}}$ matrix is carried out as explained above. The difference between smoothed FPCA via P-splines and non-penalized FPCA is only the way of computing the basis coefficients (rows of matrix A), with or without penalty, respectively. Then, an optimum set of principal components is selected and the FPCLOR is carried out. In this case, $F = \Psi_{p \times p}^{-\frac{1}{2}} G_{p \times n}$, where G is the matrix whose columns are the eigenvectors of the sample covariance matrix of $A\Psi^{1/2}$, with A the basis coefficients matrix estimated with P-splines penalty. In this method, γ is estimated via maximum likelihood without penalty.

The optimum number of principal components is chosen by GCV (see Sect. 4.3 for more details).

3.3 Method III: FPCLOR on P-spline smoothing of the principal components

In this section, we propose obtaining the principal components by maximizing a penalized sample variance that introduces a discrete penalty in the orthonormality constraint between weight principal component functions.

Taking into account the basis expansion of the sample paths given by (3), the principal component weight function f_j admits the basis expansion $f_j(t) = \sum_{k=1}^{p} b_{jk} \phi_k(t)$ and $var[\int x_i(t) f(t) dt] = b' \Psi V \Psi b$, with b being the vector of basis coefficients of the weight functions, Ψ the matrix of inner products between basis functions, and $V = n^{-1} A'A$, where $A = (a_{ij})_{n \times p}$ is the matrix of basis coefficients of the sample paths.

The ith principal component is now defined as in Eq. (9) and the basis coefficients of the factor loading f_j are obtained by solving

$$\begin{cases} \max_b \frac{b' \Psi V \Psi b}{b'(\Psi + \lambda P_d) b} \\ \text{s.t. } \|f\|^2 = b' \Psi b = 1 \quad \text{and} \quad b' \Psi b_l + b' P_d b_l = 0, \ \ell = 1, \ldots, j-1, \end{cases}$$

where λ is the smoothing parameter estimated by leave-one-out cross validation (see Sect. 4.2) and P_d the penalty matrix defined in Sect. 2.1.

Then, this variance maximization problem is converted into an eigenvalue problem, $\Psi V \Psi b = \delta(\Psi + \lambda P_d)b$, so that, applying the Cholesky factorization $LL' = \Psi + \lambda P_d$, the P-spline smoothing of FPCA turns into a classical PCA of the matrix $A\Psi(L^{-1})'$.

Finally, we carry out the FPCLOR on an optimum set of principal components obtained by the P-spline smoothing of FPCA. Then, the estimated vector β of basis coefficients of the functional parameter is given by $\hat{\beta} = F\hat{\gamma} = (L^{-1})'G\hat{\gamma}$, where G is the matrix of eigenvectors of the sample covariance matrix of $A\Psi(L^{-1})'$ and γ is estimated by the maximum likelihood criterion without penalty. The optimum number of principal components to be included in the model as regressors is chosen by GCV (see Sect. 4.3).

3.4 Method IV: FPCLOR with P-spline penalty in the maximum likelihood estimation

As developed in Reiss and Ogden (2007) for the functional linear model, we propose a smoothed version of FPCLOR that uses B-splines and roughness penalty in the regression. This penalized regression version of FPCLOR incorporates a penalty in the maximum likelihood estimation.

Taking into account the FLM in terms of non-penalized principal components and Eq. (3), the estimator of the basis coefficients of the functional parameter corresponds to $\hat{\beta} = F\hat{\gamma}$, where F is exactly the same as in Sect. 3.1 and γ is estimated by means of penalized likelihood.

Now the design matrix corresponds to $X = (\mathbf{1}|\Gamma)$, where $\Gamma = (\xi_{ij})_{n \times q}$ is a matrix of an optimal set of q functional principal components of the sample paths. Then, the penalized log-likelihood of the functional principal components logit model (4) is given by

$$\mathcal{L}^*(\lambda, \gamma) = \mathcal{L}(\gamma) - \frac{\lambda}{2}\gamma' P_d \gamma,$$

with $\gamma = (\gamma_1, \ldots, \gamma_q)'$ being the vector of the regression coefficients, P_d the penalty matrix defined in Sect. 2.1, with dimension $(q \times q)$ in this case, and $\mathcal{L}(\gamma)$ given by Eq. (5).

The optimal number of principal components and the smoothing parameter are chosen by a double GCV procedure (see Sect. 4.4 for more details).

4 Model selection

Penalized FPCLOR requires selecting an optimal number q of functional principal components and the smoothing parameter λ. Using P-spline smoothing of FPCA, the problems of high dimension, multicollinearity, and roughness in the covariables are solved. As cited in Reiss and Ogden (2007), and according to Marx and Eilers (1999) and Cardot et al. (2003), it is often assumed that the number of basis functions considered for computing P-splines has little impact as long as there are sufficiently

many knots to capture the variation in the functional parameter. Methods based on smoothed FPCA (Methods II and III) select λ in a previous step to the selection of the number q of principal components.

On the other hand, when the smoothing is applied in the likelihood estimation of the functional parameter coefficients (Methods IV and V), the sample paths are approximated by regression splines. It is known that regression splines do not control the degree of smoothness in the curves. Therefore, the selection of the number of predictor variables (non-penalized principal components for Method IV and basis functions for Method V) is essential. The optimal number of predictors and the smoothing parameter are selected in these cases by a double-GCV procedure.

4.1 Choosing λ in Method II

For Method II (Sect. 3.2) the smoothing parameter λ was selected prior to the regression. In order to select the same smoothing parameter for the n fitted P-splines, a leave-one-out cross-validation (CV) method based on minimizing the mean of the cross-validation errors over all P-splines is applied in this paper. This CV criterion consists of selecting the smoothing parameter λ that minimizes the expression

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\sum_{k=0}^{m_i} (x_{ik} - \hat{x}_{ik}^{(-k)})^2/(m_i + 1)},$$

where $\hat{x}_{ik}^{(-k)}$ are the values of the ith sample path estimated at the time t_{ik} avoiding the kth observation knot in the iterative estimation process. The number of observation knots of the ith sample path corresponds to $m_i + 1$.

4.2 Choosing λ in Method III

As in the previous section, selecting a suitable smoothing parameter is very important to control the smoothness of the weight function associated with each principal component. In this paper, CV (leave-one-out) method described in Ramsay and Silverman (2005) has been adapted by considering the discrete roughness penalty based on P-splines. It consists of selecting the value of λ that minimizes

$$CV(\lambda) = \frac{1}{p} \sum_{q=1}^{p} CV_q(\lambda),$$

where

$$CV_q(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \left\| x_i - x_i^{q(-i)} \right\|^2,$$

with $x_i^{q(-i)} = \sum_{\ell=1}^{q} \xi_{i\ell}^{(-i)} f_\ell^{(-i)}$ being the reconstruction of the sample curve x_i in terms of the first q principal components estimated from the sample of size $n-1$ that includes all sample curves except x_i.

4.3 Choosing the number of principal components in Methods I, II, and III

The optimal number q of functional principal components for Methods I, II, and III is chosen by the GCV procedure following the notes given in Craven and Wahba (1979) and Ramsay and Silverman (2005). The objective is to minimize

$$\text{GCV}(q) = \left(\frac{n}{n - \text{tr}(H^q)} \right) \left(\frac{\text{MSE}(q)}{n - \text{tr}(H^q)} \right), \tag{10}$$

where $\text{MSE}(q) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i^q)^2$ and H^q is the "hat" matrix given by

$$H^q = W_q^{1/2} X \left(X' W_q X \right)^{-1} X' W_q^{1/2},$$

with $W_q = \text{Diag}[\hat{\pi}_i^q (1 - \hat{\pi}_i^q)]$ as the weight matrix. The design matrix X depends on the considered method as follows:

Method I: $X = (1|\Gamma)$, with Γ being the matrix comprising the columns of the first q functional principal components of $A\Psi^{1/2}$, with A the matrix of basis coefficients of the sample paths estimated via regression splines, and $\Psi^{1/2}$ the square root of the matrix of the inner products between B-spline basis functions.

Method II: $X = (1|\Gamma)$, with Γ being the matrix comprising the columns of the first q functional principal components of $A\Psi^{1/2}$, with A the basis coefficients estimated via penalized splines (P-splines).

Method III: $X = (1|\Gamma)$, with Γ being the matrix comprising the columns of the first q functional principal components of $A\Psi(L^{-1})'$, with A the matrix of basis coefficients of the sample paths estimated via regression splines, and L given by the Cholesky decomposition.

4.4 Choosing the number of predictors and the smoothing parameter in Methods IV and V

In Methods IV and V, the log-likelihood is penalized and the parameters of the model are simultaneously chosen by a double-GCV. In Method IV, the double-GCV consists in computing the GCV error (10) for each number of principal components q and each λ of a grid of possible values. Then, q is selected by minimizing the mean of the GCV error over all possible values of λ. Once q is selected, the value of λ with the lowest GCV error is chosen. In Method V, the procedure is the same by replacing the number q of principal components by the number p of basis functions.

The design matrix X of Methods IV and V corresponds to

Method IV: $X = (1|\Gamma)$, with Γ being the matrix comprising the columns of the q first functional principal components of $A\Psi^{1/2}$ and A the matrix of basis coefficients of the sample paths estimated via regression splines.

Method V: $X = (1|A\Psi)$, with A being the matrix of basis coefficients of the sample paths approximated by regression splines.

When the log-likelihood criterion is penalized by using P-splines, the "hat" matrix is given by

$$H = W^{1/2} X \left(X' W X + \lambda P_d \right)^{-1} X' W^{1/2},$$

with P_d defined as in Sect. 2.1.

5 Simulation study

The good performance of the proposed penalized estimation approaches to estimate the parameter function and to predict the response is evaluated in this section on two different simulation schemes, and the results compared with the ones provided by non-penalized FPCLOR (Method I).

On the other hand, the ability of the proposed approaches to forecast a binary response and classify a set of curves has also been compared with a competitive classification procedure as the partial least squares approach for functional linear discriminant analysis (FLD-PLS) introduced by Preda et al. (2007) and its basis expansion estimation with B-spline basis proposed in Aguilera et al. (2010). It is important to clarify that we can compare our results with the prediction errors and classification rates given by this procedure but the estimated parameter functions are not comparable because they correspond to different regression models from a theorist point of view.

Functional linear discriminant analysis is used to classify a set of sample curves in the two groups defined by a binary response. Taking into account the equivalence between linear discriminant analysis and linear regression, it is known that the discriminant function is the functional parameter associated with the functional linear regression of Y on $\{X(t) : t \in T\}$ with Y recoded as

$$Y = \begin{cases} -\sqrt{p_0/p_1} & \text{if } Y = 1, \\ \sqrt{p_1/p_0} & \text{if } Y = 0, \end{cases}$$

where $p_0 = P[Y = 0]$ and $p_1 = P[Y = 1]$. The FLD-PLS approach is based on using functional partial least squares to estimate the functional parameter associated with this functional linear model. For a detailed study of this classification procedure, the interested lector is referred to Preda et al. (2007).

5.1 Case I: simulation of waveform data

This data set was introduced by Breiman et al. (1984) and used later by Hastie et al. (1994), Ferraty and Vieu (2003), and Escabias et al. (2007). Following the simulation scheme developed in Escabias et al. (2007), 1000 curves of two different classes of sample curves were simulated with 500 curves for each one according to the random functions

$$x(t) = uh_1(t) + (1 - u)h_2(t) + \varepsilon(t) \quad \text{(class 1)},$$
$$x(t) = uh_1(t) + (1 - u)h_3(t) + \varepsilon(t) \quad \text{(class 2)},$$

with u and $\varepsilon(t)$ being uniform and standard normal simulated random variables, respectively, and

$$h_1(t) = \max\{6 - |t - 11|, 0\}, \quad h_2(t) = h_1(t - 4), \quad h_3(t) = h_1(t + 4).$$

Each sample curve was simulated at 101 equally spaced points in the interval $[1, 21]$. An example of simulated sample paths for class 1 (a) and class 2 (b) is shown in Fig. 1. The binary response variable was defined as $Y = 0$ for the curves of the first
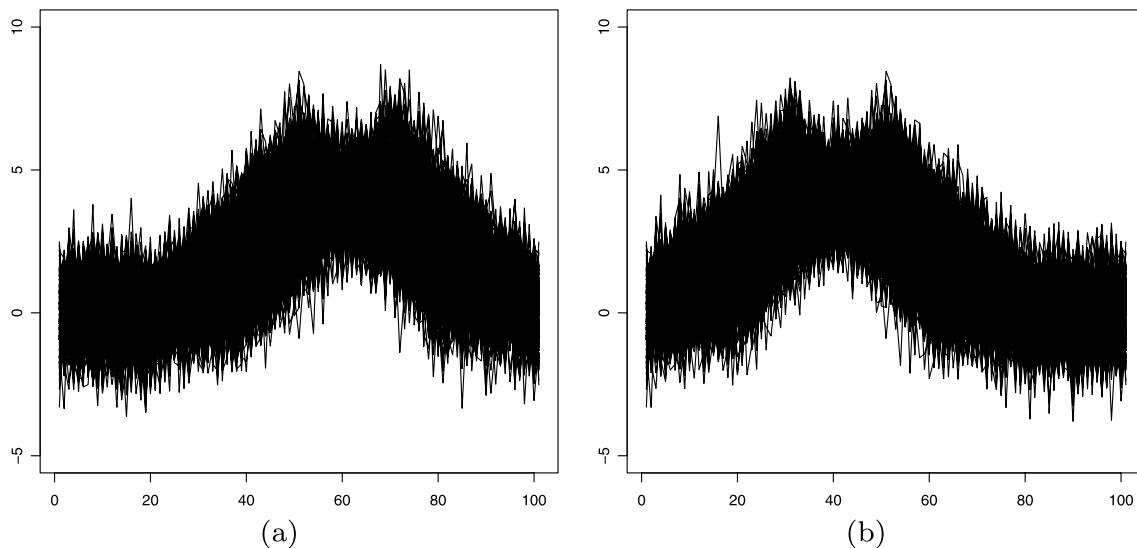
Fig. 1 Case I. Simulated sample curves for class 1 (**a**) and class 2 (**b**) in one of the 100 simulations

Fig. 2 Case I. Estimated functional parameter for one of the 100 simulations. The functional parameter is estimated by Method I (*black short dashed line*), Method II (*red solid line*), Method III (*blue dotted line*), Method IV (*green dashed* and *dotted line*), and Method V (*pink large dashed line*) (Color figure online)
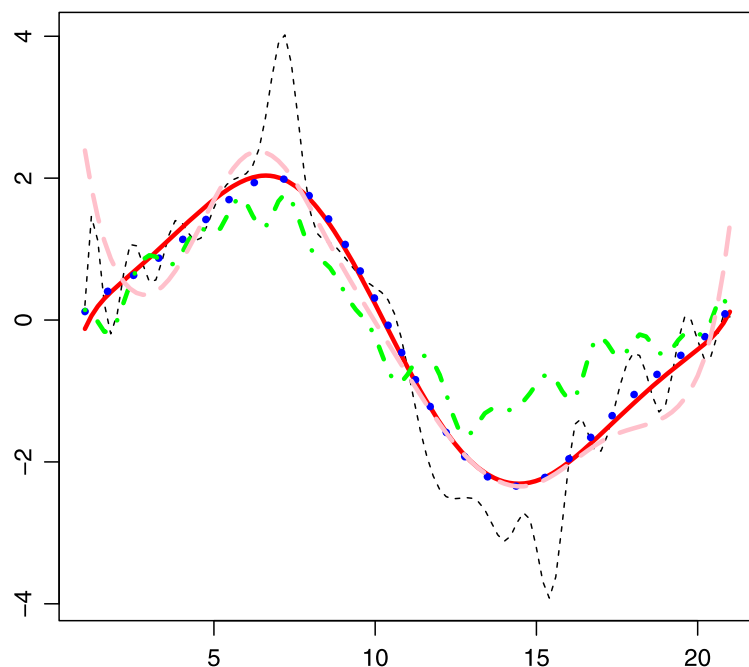
class and $Y = 1$ for the ones of the second class. After simulating the data, least squares approximation (with and without penalty) in terms of the cubic B-spline functions defined on 30 equally spaced knots in the interval $[1, 21]$ was performed for each sample curve. When working with P-splines, the number of basis knots is not so critical and only a large number of equally spaced knots is needed. The choice of the P-splines parameters was discussed by Eilers and Marx (1996), Ruppert (2002), and Currie and Durban (2002). For the case of equally spaced observations, they conclude that using one knot for every four or five observations up to a maximum of 40 knots is often sufficient.

In order to corroborate the good performance of the penalized estimation approaches proposed in this paper, 100 repetitions of this simulation scheme were carried out. The functional parameter estimated by means of the five different methods presented in previous sections are displayed in Fig. 2 for one of the simulations. The

Table 1 Case I. Mean and standard deviation (S.D.) for the GCV errors of the models estimated by Methods I, II, III, IV, and V

Method	GCV error	
	Mean	S.D.
Method I	0.00003	0.000008
Method II	0.00003	0.000007
Method III	0.00002	0.000007
Method IV	0.00014	0.000112
Method V	0.00022	0.000120

mean of the estimated functional parameters over the 100 simulations is plotted in Fig. 3 for each of the five estimation approaches and for FLD-PLS next to confidence bands computed as the mean ± 2 the standard deviation. The functional parameter estimated by FLD-PLS (discriminant function) is not comparable with the others because is associated with different regression models.

Let us observe that there are important differences between the estimations provided by the non-penalized FPCLOR approach (Method I) and the other four methods based on penalized estimation of the FLM. The functional parameter estimated by non-penalized FPCLOR (Method I) is not smooth and affected by high variability. It is therefore difficult to interpret and needs to be smoothed. The estimations provided by Methods II, III, and V are quite similar, but sometimes the estimations provided by Method V are over-smoothed and lose the control in the extremes of the observation interval. On the other hand, when the P-spline penalty is introduced in the log-likelihood criterion of a FPCLOR model (Method IV), the estimated functional parameter is smoother than the one given by Method I, but it is not smooth enough and is affected by some variability. Therefore, the necessity of using smoothed functional principal components as explicative variables is obvious. The best estimations are achieved with Methods II and III, providing the smoothest parameter functions with the least variability.

In order to compare the goodness of fit and the forecasting ability of the five estimation approaches the box-plots related to the area under ROC curve and the MSE distributions (on 100 test samples) are shown in Fig. 4. It can be observed that Methods I, II, and III based on non-penalized principal component logit regression result in much more accurate predictions than Methods IV and V based on penalized likelihood estimation. Among them, Method II achieves the highest area under ROC curve and Method III the smallest MSE and GCV error (see Table 1 for GCV errors). Let us observe that the FLD-PLS approach gets the highest prediction error, but has good classification ability similar to Methods II and III.

5.2 Case II: simulation of the Ornstein–Uhlenbeck process

In order to obtain more general conclusions about the behavior of the proposed methods, a second simulation study where the functional parameter is known has been developed.

Let us consider $\{O_t : t \in [0, T]\}$ the well known zero mean Gaussian process known as the Ornstein–Uhlenbeck process. The simulated sample paths were com-
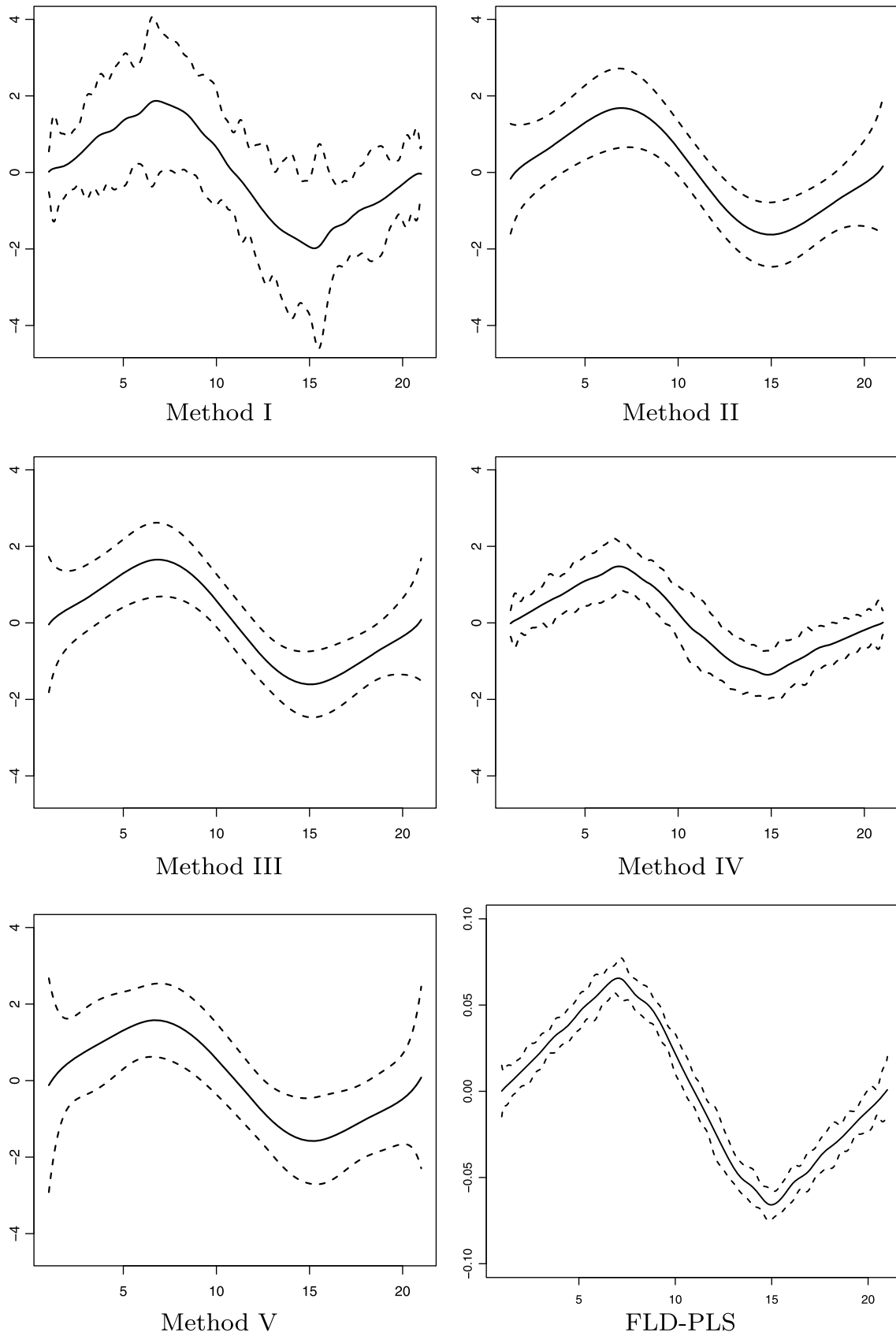
Fig. 3 Case I. Mean of the functional parameters and the confidence bands (computed as the mean ±2 the standard deviation) estimated by Methods I, II, III, IV, V and FLD-PLS over the 100 simulations
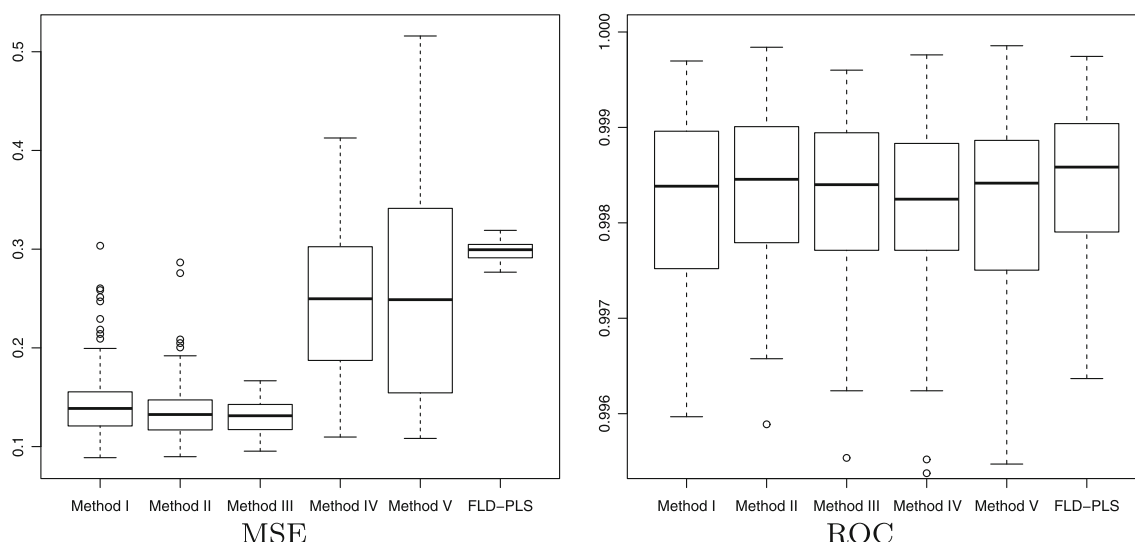
Fig. 4 Case I. Area under ROC curve and MSE distribution (for the test samples of the 100 repetitions) given by Methods I, II, III, IV, V and FLD-PLS

puted taking into account the decomposition of this process in terms of principal components truncated at the 14th term

$$O^{14}(t) = \sum_{i=1}^{14} \lambda_i f_i(t)\xi_i,$$

with λ_i and f_i being the eigenvalues and eigenfunctions associated with the covariance function given by $C(t, s) = P \exp(-\alpha|t - s|)$, and ξ_i being the corresponding principal components that have distribution $N(0, 1)$. This principal component reconstruction is a smooth version of the Ornstein–Uhlenbeck process that explains 99.4 % of its total variance.

In order to have noisy observations, a random error $\varepsilon(t)$ with distribution $N(0, \sigma^2)$ was added so that the simulated process is given by

$$X(t) = O^{14}(t) + \varepsilon(t). \tag{11}$$

The variance of the errors σ^2 was chosen by controlling $R^2 = Var[O^{14}]/Var[X]$ close to 0.8. The parameters used for the simulation were $T = 4$, $P = 1$, and $\alpha = 0.1$. In this study, 200 samples of 100 and 50 sample curves of the contaminated process given by Eq. (11) were simulated for training and test samples, respectively, at 41 equally spaced knots in the interval [0, 4].

In order to simulate the binary response associated with each sample path x_i, we have considered the parameter function

$$\beta(t) = 6\cos(0.25\pi t) - 0.5\sin(0.25\pi t)$$

and computed the expectations π_i according to Eq. (1). Then, the associated response value y_i was simulated by a Bernoulli distribution with parameter π_i.

Let us remember that the main purpose of this work is to improve the estimation of the functional parameter in functional logit regression, providing in addition a good classification rate. In order to check the ability of the proposed penalized spline approaches to estimate the functional parameter of the logit model provided by the six

methods, the mean of the estimated functional parameters over the 200 simulations is plotted in Fig. 5 next to the original parameter function and the confidence bands (computed as the mean ± 2 the standard deviation). The integrated mean squared error with respect to the original functional parameter was also computed for each method by using the following expression:

$$\text{IMSE}\beta = \left(\frac{1}{T} \int_T \left(\beta(t) - \hat{\beta(t)} \right)^2 \right)^{1/2}.$$

The box plots with the distribution of the IMSEβ for the five estimation approaches of the functional parameter associated with the logit model are displayed in Fig. 6. The means and standard deviations of these errors appear in Table 2.

Let us observe that Methods I (non-penalized FPCLOR approach) and IV provide the least smooth estimates with the worst results given by Method IV that is affected by high variability. On the other hand, Methods II and III provide again similar results with smoother estimates affected by high variability in the extremes of the observation interval. By observing the estimated mean functions, it can be observed again that the estimations provided by Method V are over-smoothed and have less variability than those by Methods II and III. The integrated errors with respect to the original parameter function are also higher for Method V than for Methods II and III. The discriminant function associated with the FLD-PLS approach is noisy and affected by a very high degree of variability (see Fig. 7).

The forecasting performance and classification ability of the six methods can be tested by comparing the distributions of the mean squared error (MSE) and ROC area displayed in Fig. 8. According to the MSE, Methods III and V are quite similar, providing the smallest prediction errors, while FLD-PLS gives the highest prediction errors. With respect to the ROC area, Methods III and V achieve also the highest values followed by Method II and FLD-PLS. On the other hand, Method IV provides the worst classification performance (smallest area under the ROC curve), although in all cases the ability of the considered methods to classify the curves is very good with a median greater than 93 %. From this simulation, we can conclude that Method III provides an accurate estimation of the functional parameter and has the best classification ability followed by Methods V and II that give similar results. In addition, Method III outperforms competitive methods such as FLD-PLS in both predictive and classification ability.

5.3 Computational considerations

All process have been analyzed in the same experimental conditions (computers with the same conditions). Specifically, the simulations are run on a cluster of 30 blade servers each one with two Intel XEON $E5420$ processors running at 2.5 GHz and with 16GB of RAM memory. Each processor has four cores, and the experiments are carried out on virtualized Windows XP machines, each one with one virtualized processor and 1GB of RAM memory. Twenty-five Windows XP systems have been simultaneously used to carry out the simulation procedures.

The numerical results presented in this paper have been obtained by using the R project software for statistical computing (http://www.r-project.org/). The authors

Fig. 5 Case II. Mean of the functional parameters and the confidence bands (computed as the mean ±2 the standard deviation) (at *left*) and the true functional parameter (*solid line*) superposed with the mean of the estimated functional parameters (*dashed line*) (at *right*) provided by Methods I, II, III, IV, and V over 200 simulations
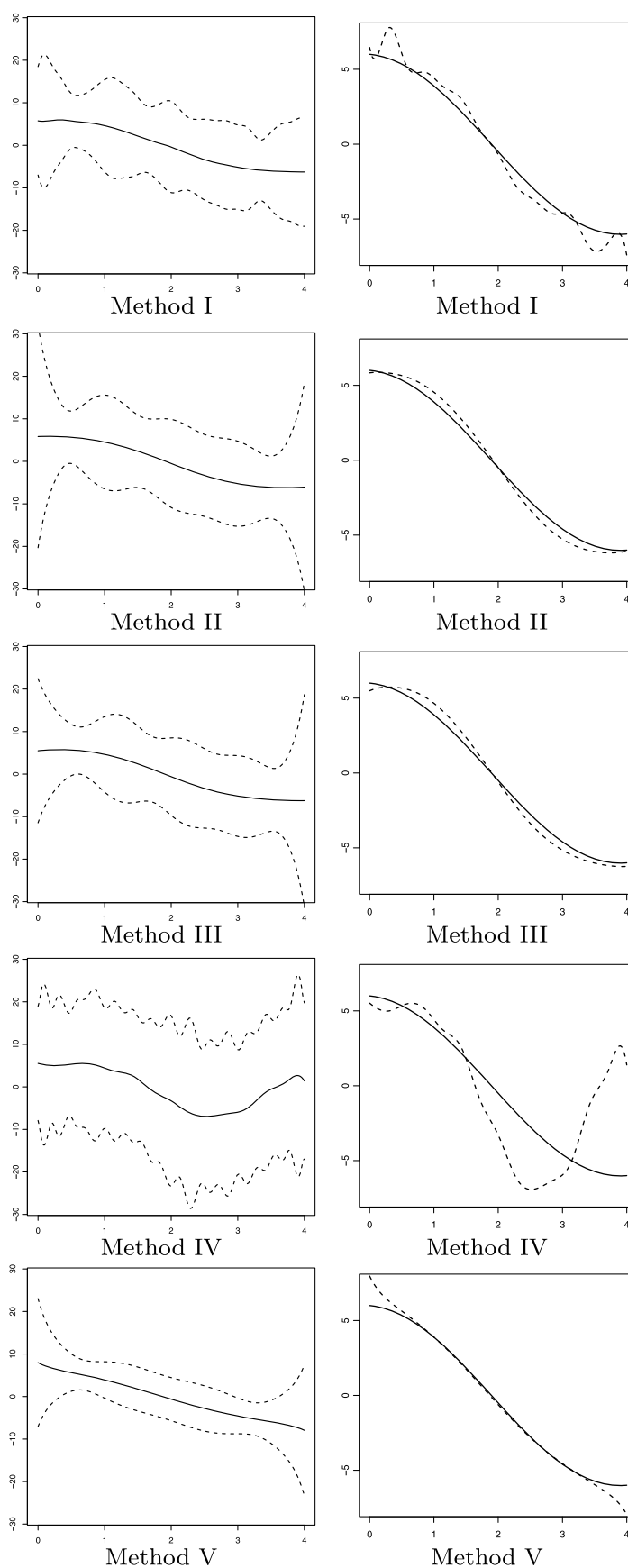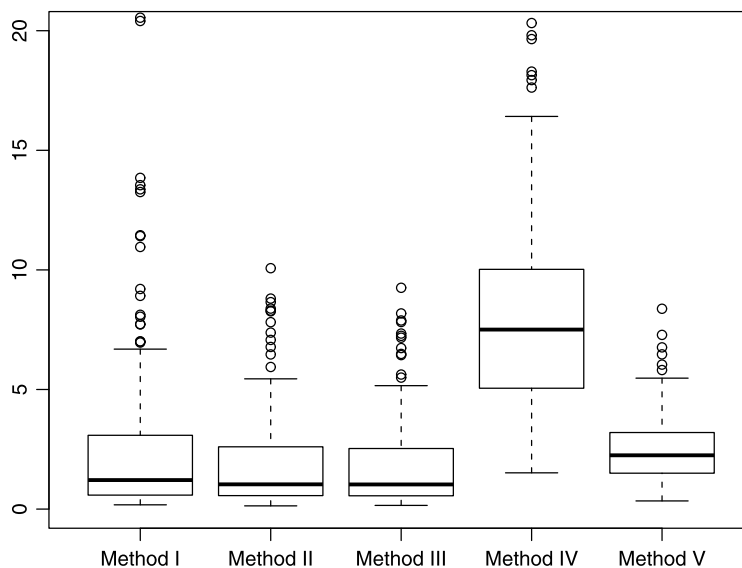
Fig. 6 Case II. Box-plot of the distribution of the IMSEβ for the estimated parameter functions on 200 repetitions given by Methods I, II, III, IV, and V

Fig. 7 Case II. Mean of the functional parameters and the confidence bands (computed as the mean ± 2 the standard deviation) estimated by FLD-PLS method over 200 simulations
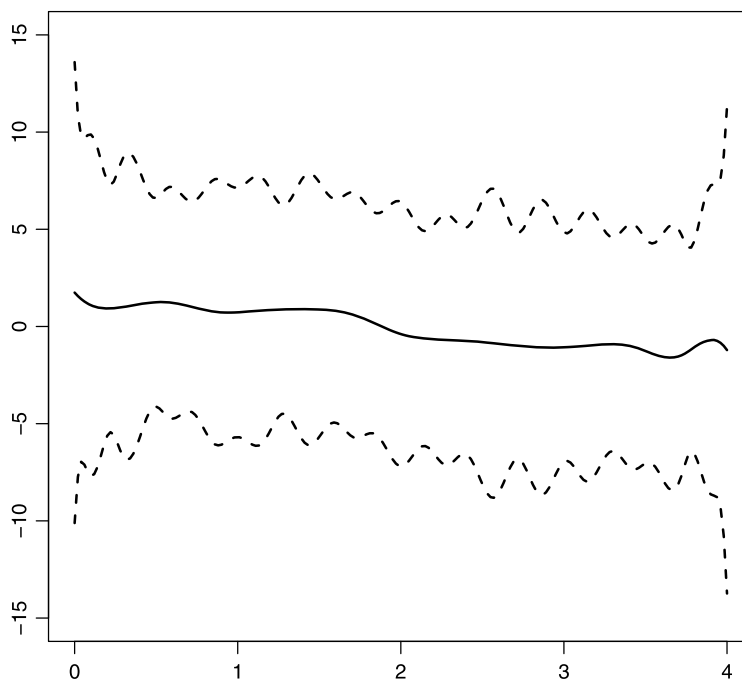
Table 2 Case II. Mean and standard deviation of the IMSEβ

	Method I	Method II	Method III	Method IV	Method V
Mean	3.1893	1.8931	1.8166	8.0691	2.5332
SD	9.4394	2.0289	1.8645	4.1827	1.3898

have developed their own scripts with R code by using specific functions of the R packages: fda, stats, design, and plsr. When functions for estimating the proposed penalized spline methods had not been available in the R packages, the authors have developed original R functions.
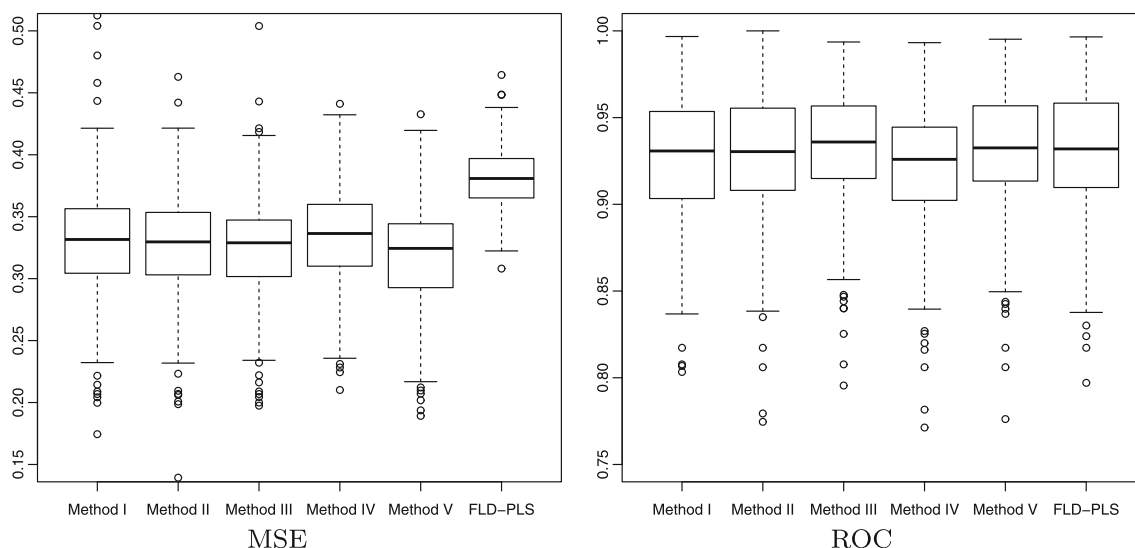
Fig. 8 Case II. Area under ROC curve and MSE distribution (for the test samples of the 200 repetitions) given by Methods I, II, III, IV, V and FLD-PLS

6 Real data application

The good performance of Methods II and III based on smoothed functional principal component analysis was proved in Sect. 5. These methods are now compared with non-penalized FPCLOR on a real data set based on the biscuit productions of the manufacturer Danone.

The manufacturer Danone aims to use only flour that guarantees good product quality. The quality of a biscuit depends on the quality of the flour used to make it. In this paper, smoothed functional principal component logit regression models are applied in order to estimate the quality of the biscuits.

There are several kinds of flour that are distinguished by their composition. The quality of cookies made with each flour can be good or bad. The aim is to classify a cookie as good or bad from the resistance (density) of the dough observed in a certain interval of time during the kneading process. To solve this problem, a functional logit model is used to estimate the quality (Y) that takes value $Y = 1$ if the quality of a cookie is good and $Y = 0$ if bad. For a given flour, the resistance of dough is recorded every two seconds during the first 480 seconds of the kneading process. Thus, the predictor data set is given by a set of curves $\{X(t_i), i = 0, \ldots, 240\}$ observed at 241 equally spaced time points in the interval $[0, 480]$. After kneading, the dough is processed to obtain biscuits. The sample consists of 90 different flours whose curves of resistance during the kneading process are considered independent realizations of the continuous stochastic process $\mathbf{X} = \{X(t) : t \in [0, 480]\}$. Of the 90 observations, we have 40 for $Y = 0$ (bad cookies) and 50 for $Y = 1$ (good cookies). The sample paths are displayed in Fig. 9. Different versions of functional discriminant analysis based on functional PLS regression were applied to classify these curves in Preda et al. (2007) and Aguilera et al. (2010).

The sample of 90 flours is randomly divided into a training sample of size 60 (35 good and 25 bad) and a test sample of size 30 with the same number of curves for the two classes. In order to obtain more general conclusions, 100 different random divisions of the sample of 90 flours into training and test samples were considered. Taking
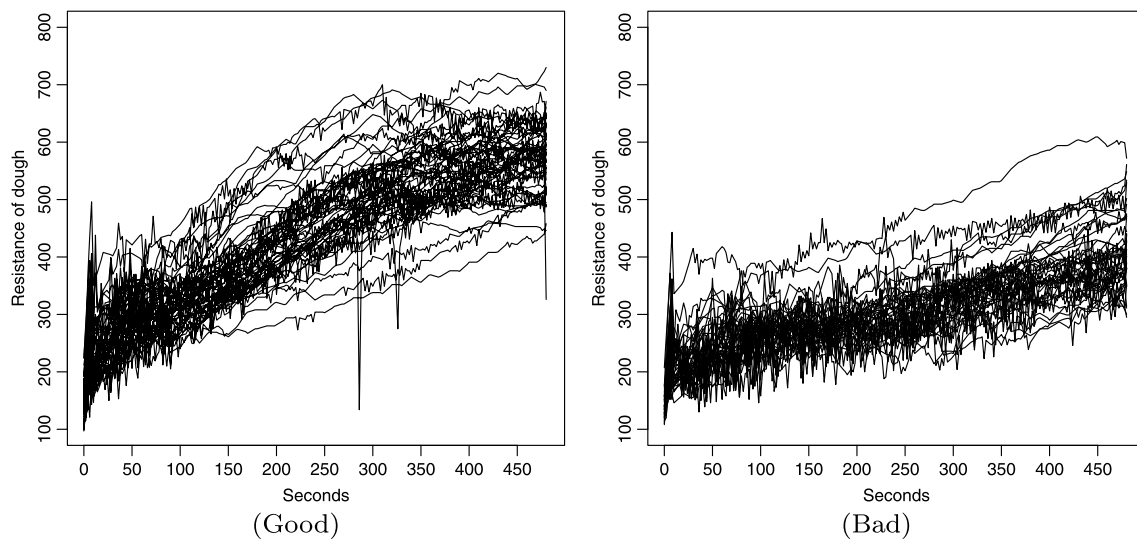
Fig. 9 Danone data set. Sample curves of resistance of dough for good (*left*) and bad (*right*) flours
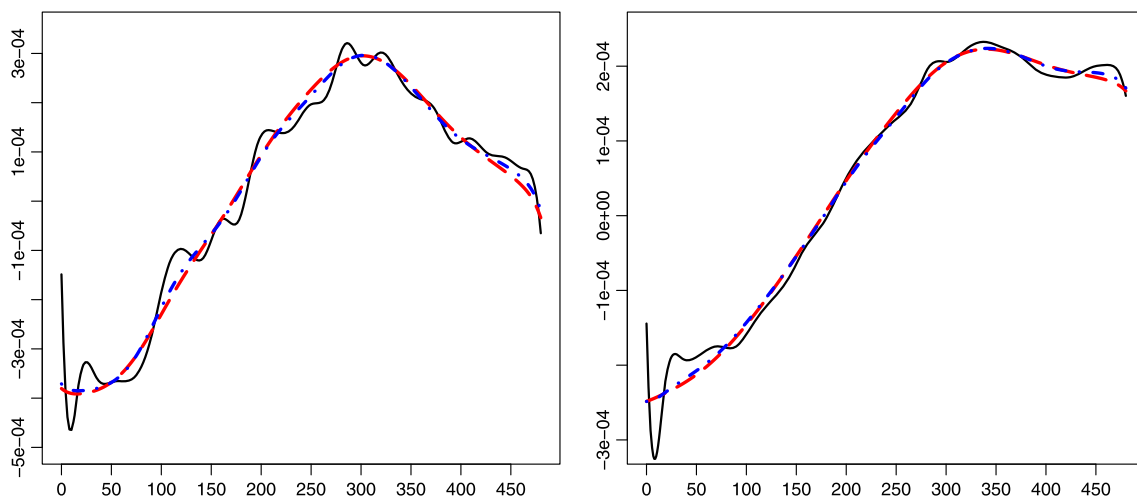

Fig. 10 Danone data set. Functional parameter estimated from one of the training samples (*left*) and the mean of the functional parameters estimated from the 100 training samples (*right*) by using Method I (*black solid line*), Method II (*red long dashed line*), and Method III (*blue dashed and dotted line*) (Color figure online)

into account that the resistance of dough is a smooth curve measured with error, the reconstruction of the true functional form was carried out by using a cubic B-splines basis defined on 28 equally spaced knots in the interval $[0, 480]$. Once the sample curves were approximated, P-spline smoothed functional principal component logit regression was performed (Methods II and III). Both methods were compared with the well known non-penalized functional principal components logit regression proposed in Escabias et al. (2004). The estimated functional parameter from one of the training samples is shown in Fig. 10. The mean of the functional parameters estimated by Methods I, II, and III for the 100 different divisions of the sample in training and test samples are displayed in Fig. 11 next to confidence bands computed as the mean ± 2 the standard deviation.

It is clearly observed that the functional parameter estimated by non-penalized FPCLOR (Method I) is not smooth. On the other hand, methods based on smoothed
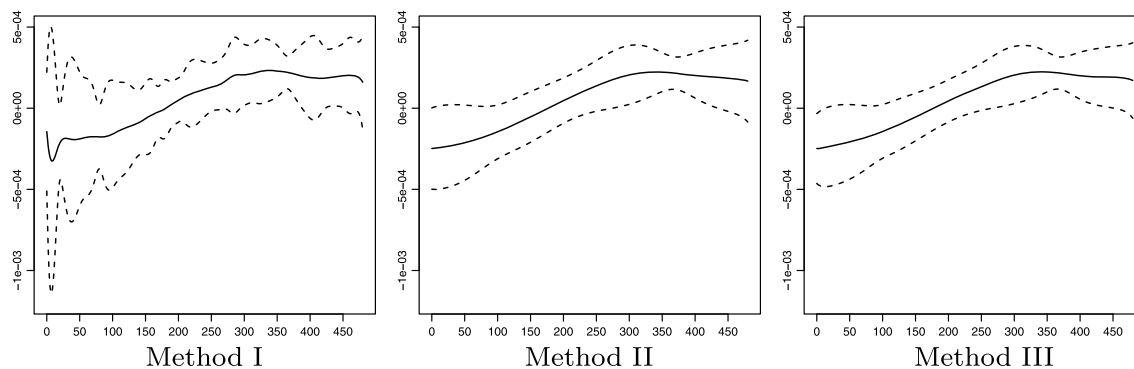
Fig. 11 Danone data set. Mean of the functional parameters over 100 training samples and confidence bands (computed as the mean ±2 the standard deviation) estimated for Methods I, II, and III

Fig. 12 Danone data set. Mean of the functional parameters and the confidence bands (computed as the mean ±2 the standard deviation) estimated by FLD-PLS method, over the 100 simulations
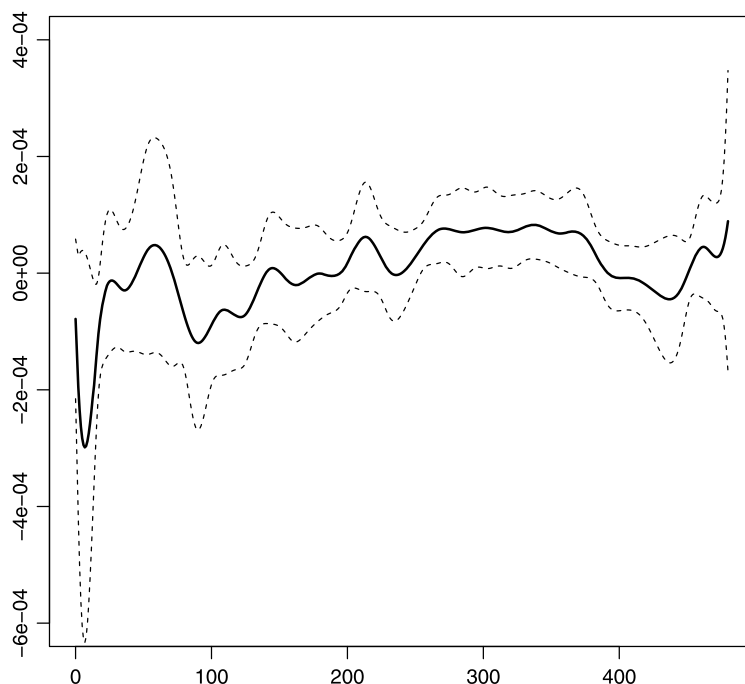
functional principal components (Methods II and III) control the roughness of the estimated curve, providing an accurate and smooth estimation of the parameter function. From Fig. 12, it can be seen that the functional parameter associated with the FLD-PLS approach is not smooth at all. Let us remember that this curve is not comparable to the functional parameter of the logit model displayed in Fig. 11.

In order to check the ability of P-spline smoothing FPCLOR approaches to forecast the binary response and to classify the curves as good or bad, the box-plots of the distribution of area under ROC curve, MSE and CCR with cutpoint 0.5 (for the 100 test samples) are shown in Fig. 13. The mean and standard deviation of the misclassification errors with cut point 0.5 can be seen in Table 3. It can be observed that the penalized spline approaches provide smaller prediction errors than the non-penalized FPCLOR and FLD-PLS. On the other hand, the area under the ROC curve is higher for FLD-PLS, although the differences between the four methods are not significant and all give very high ROC area with median greater than 97 %. By comparing Methods II and III based on P-spline smoothing of functional PCA, it can be concluded
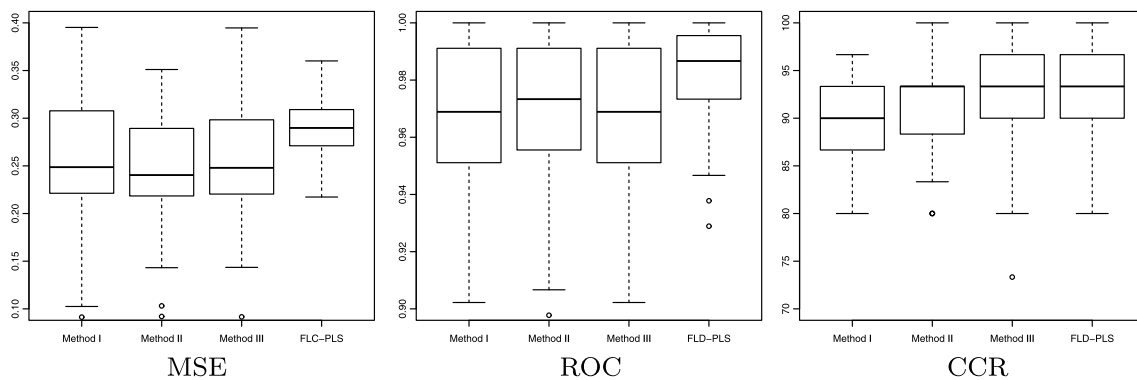
Fig. 13 Danone data set. MSE, area under ROC curve and CCR distributions (for the 100 test samples) by using Methods I, II, and III

Table 3 Danone data set. Misclassification rate averaged over 100 test samples

	Method I	Method II	Method III	FLD-PLS
Mean	0.1626	0.1242	0.0885	0.0798
SD	0.2369	0.1861	0.0915	0.0406

that both achieve similar results with Method II providing the highest area under ROC curve and the smallest MSE. With respect to the misclassification errors, Method III and LDA-PLS gives the best performance.

7 Conclusions

In order to solve the problem of multicollinearity in functional logit regression and to control de smoothness of the functional parameter estimated from noisy smooth sample curves, four different penalized spline (P-spline) estimations of the functional logit model are proposed in this paper. Let us take into account that the aim of logit model is not only to classify a set of curves into two groups but mainly to interpret the relationship between the binary response and the functional predictor in terms of the functional parameter. Because of this, our main purpose is to improve the estimation of the functional parameter of a functional logit model, providing in addition a good classification rate.

A P-spline penalty measures the roughness of a curve in terms of differences of order d between coefficients of adjacent B-spline basis functions. The proposed smoothing approaches are based on B-spline expansion of the sample curves and the parameter function, and P-spline estimation of the functional parameter. The difference is in how to introduce the penalty in the model. Three of the considered approaches (Methods II, III, and IV) are based on functional principal component logit regression that consists in regressing the binary response on a reduced set of functional principal components. In Method II, the P-spline penalty is introduced by performing the functional PCA on the P-spline least squares approximation of the sample curves from discrete observations. Method III introduces the P-spline penalty in the own formulation of functional PCA and the principal components are computed by maximizing a penalized sample variance that introduces a discrete penalty

in the orthonormality constraint between the principal components weight functions. In Method IV, the P-spline penalty is used in the maximum likelihood estimation of the functional parameter in terms of functional principal components. On the other hand, direct P-spline likelihood estimation in terms of B-spline functions is also considered (Method V).

Two simulation studies and an application with real data were performed to test the ability of the proposed P-spline smoothing approaches to provide an accurate and smooth estimation of the functional parameter and a good classification performance. Leave-one-out cross-validation and generalized cross-validation are adapted to select the different parameters (smoothing parameter and number of principal components or basis functions) associated with the considered approaches. In the case of the P-spline approximation of the sample curves from equally spaced observations, a relatively large number of equally spaced basis knots is a good choice for the definition of the B-spline basis. The results provided by the different smoothing approaches are compared with the estimations provided by non-penalized FPCLOR on least squares approximation of sample curves with B-spline basis (Method I) and by the partial least squares estimation approach for functional linear discriminant analysis with (FLD-PLS).

From the simulation study and the real data application, it can be concluded that the estimation of the functional parameter given by the P-spline approaches is much smoother than the one given by the non-penalized FPCLOR, although in some cases Method IV gives worse results. In fact, Methods I and IV provide non-smooth estimations affected by high variability. The most accurate and smoothest estimations of the parameter function are provided by Methods II and III, based on P-spline estimation of functional PCA with B-spline basis. On the other hand, the estimations given by Method V are less accurate and oversmoothed. In relation to the forecasting ability of the proposed methodologies, Methods II and III provide the least prediction errors, followed by Method V that also gives accurate results. The classification performance of all the methods is very good, with Methods II, III, and V being the most competitive. On the other hand, the LDA-PLS approach gives very high classification rates similar to Methods II, III, and V, but its forecasting errors are much higher.

In summary, it can be concluded that the penalized approaches represented by Methods II and III are preferred because they provide the most accurate estimation of the parameter function and have the best forecasting and classification performance, with Method II having lower computational cost.

An intuitive explanation of the fact that Methods II and III provide better estimation of the functional parameter could be that these approaches develop a penalized smoothing of the sample curves before estimating the regression model and select the smoothing parameter according to the mean squared error with respect to the observed sample curves. This way, the smoothing of the curves provides smoothed estimation of the functional parameter. On the other hand, the results given by Method V are not so good because the roughness of the functional parameter is directly penalized in the ML estimation but the smoothing parameter is selected by minimizing the prediction error without taking into account the smoothness of the sample curves. Finally, Method IV gives the worst estimations of the functional parameter because this approach does not penalize the roughness of any of the functions involved in the

analysis and the penalty is only on the regression coefficients in terms of the non-penalized principal components.

References

Agresti A (1990) Categorical data analysis. Wiley, New York

Aguilera A, Escabias M, Preda C, Saporta G (2010) Using basis expansion for estimating functional pls regression. Applications with chemometric data. Chemom Intell Lab Syst 104:289–305

Aguilera A, Escabias M, Valderrama M (2008) Discussion of different logistic models with functional data. Application to systemic lupus erythematosus. Comput Stat Data Anal 53(1):151–163

Aguilera A, Gutiérrez R, Ocaña F, Valderrama M (1995) Computational approaches to estimation in the principal component analysis of a stochastic process. Appl Stoch Models Data Anal 11(4):279–299

Aguilera A, Gutiérrez R, Ocaña F, Valderrama M (1996) Approximation of estimators in the pca of a stochastic process using B-splines. Commun Stat, Simul Comput 25(3):671–690

Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and regression trees. Wadswort

Cardot H, Crambes C, Kneip A, Sarda P (2007) Smoothing splines estimators in functional linear regression with errors-in-variables. Comput Stat Data Anal 51:4832–4848

Cardot H, Ferraty F, Sarda P (2003) Spline estimators for the functional linear model. Stat Sin 13:571–591

Cardot H, Sarda P (2005) Estimation in generalized linear models for functional data via penalized likelihood. J Multivar Anal 92(1):24–41

Craven P, Wahba G (1979) Smoothing noisy data with spline functions—estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math 31(4):377–403

Currie I, Durban M (2002) Flexible smoothing with P-splines: a unified approach. Stat Model 2(4):333–349

De Boor C (2001) A practical guide to splines, revised edn. Springer, Berlin

Eilers P, Marx B (1996) Flexible smoothing with B-splines and penalties. Stat Sci 11(2):89–121

Eilers P, Marx B (2002) Generalized linear additive smooth structures. J Comput Graph Stat 11(4):758–783

Escabias M, Aguilera A, Valderrama M (2004) Principal component estimation of functional logistic regression: discussion of two different approaches. J Nonparametr Stat 16(3–4):365–384

Escabias M, Aguilera A, Valderrama M (2005) Modeling environmental data by functional principal component logistic regression. Environmetrics 16(1):95–107

Escabias M, Aguilera A, Valderrama M (2007) Functional pls logit regression model. Comput Stat Data Anal 51(10):4891–4902

Ferraty F, Vieu P (2003) Curves discrimination: a nonparametric functional approach. Comput Stat Data Anal 44:161–173

Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer, Berlin

Harezlak J, Coull B, Laird N, Magari S, Christiani D (2007) Penalized solutions to functional regression problems. Comput Stat Data Anal 51(10):4911–4925

Hastie T, Buja A, Tibshirani R (1994) Flexible discriminant analysis by optimal scoring. J Am Stat Assoc 89(428):1255–1270

Hastie T, Tibshirani R (1990) Generalized additive models. Chapman & Hall, London

Hastie T, Tibshirani R (1993) Varying-coefficient models (with discussion). J R Stat Soc B 55:757–796

James G (2002) Generalized linear models with functional predictors. J R Stat Soc B 64(3):411–432

Le Cessie S, Van Houwelingen J (1992) Ridge estimators in logistic regression. Appl Stat 41(1):191–201

van der Linde A (2008) Variational Bayesian functional pca. Comput Stat Data Anal 53(2):517–533

Marx B, Eilers P (1999) Generalized linear regression on sampled signals and curves. A p-spline approach. Technometrics 41(1):1–13

Müller H (2005) Functional modelling and classification of longitudinal data. Board of the Foundation of Scand J Stat 32:223–240

Ocaña F, Aguilera A, Escabias M (2007) Computational considerations in functional principal component analysis. Comput Stat 22(3):449–465

Ocaña F, Aguilera A, Valderrama M (2008) Estimation of functional regression models for functional responses by wavelet approximations. In: Dabo-Niang S, Ferraty F (eds) Functional and operatorial statistics. Physica-Verlag, Heidelberg, pp 15–22

O'Sullivan F (1986) A statistical perspective on ill-posed inverse problems. Stat Sci 1:505–527

Preda C, Saporta G, Lévéder C (2007) Pls classification for functional data. Comput Stat 22:223–235

Ramsay JO, Silverman BW (2002) Applied functional data analysis: methods and case studies. Springer, Berlin

Ramsay JO, Silverman BW (2005) Functional data analysis. Springer, Berlin

Ratcliffe S, Heller G, Leader L (2002) Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression. Stat Med 21:1115–1127

Reinsch C (1967) Smoothing by spline functions. Numer Math 10:177–183

Reiss P, Ogden R (2007) Functional principal component regression and functional partial least squares. J Am Stat Assoc 102(479):984–996

Ruppert D (2002) Selecting the number of knots for penalized splines. J Comput Graph Stat 11(4):735–757