Functional PCA and Base-Line Logit Models

- Manuel Escabias; Ana M. Aguilera, M. Carmen Aguilera-Morillo

- Functional PCA and Base-Line Logit Models

- *Journal of Classification 31:296-324 (2014)*

- DOI: 10.1007/s00357-014-9162-y

Functional PCA and Base-Line Logit Models

Manuel Escabias

Universidad de Granada, Spain

Ana M. Aguilera

Universidad de Granada, Spain

M. Carmen Aguilera-Morillo

Universidad Carlos III de Madrid, Spain

Abstract: In many statistical applications data are curves measured as functions of a continuous parameter as time. Despite of their functional nature and due to discrete-time observation, these type of data are usually analyzed with multivariate statistical methods that do not take into account the high correlation between observations of a single curve at nearby time points. Functional data analysis methodologies have been developed to solve these type of problems. In order to predict the class membership (multi-category response variable) associated to an observed curve (functional data), a functional generalized logit model is proposed. Base-line category logit formulations will be considered and their estimation based on basis expansions of the sample curves of the functional predictor and parameters. Functional principal component analysis will be used to get an accurate estimation of the functional parameters and to classify sample curves in the categories of the response variable. The good performance of the proposed methodology will be studied by developing an experimental study with simulated and real data.

Keywords: Functional data analysis; Nominal logit regression; Principal components.

Authors' Addresses: M. Escabias, Facultad de Farmacia, Universidad de Granada, Campus de Cartuja 18071 Granada, Tel.: +34-958-240640, Fax: +34-358-249046, e-mail: escabias@ugr.es; A.M. Aguilera, Facultad de Ciencias, Universidad de Granada, Campus Fuentenueva 18071 Granada, e-mail: aaguiler@ugr.es; M.C. Aguilera-Morillo, Departamento de Estadística, Universidad Carlos III de Madrid, Avda. de la Universidad, 30. 28911 Leganés, e-mail: maguiler@est-econ.uc3m.es.

1. Introduction

In the last fifteen years, a new area of research in statistics has been developed, called Functional Data Analysis (FDA), that emerges as a generalization of multivariate data analysis techniques to the case where the data are curves. The interested reader can consult the books Ramsay and Silverman (2002; 2005). Applications of FDA in different fields have been recently developed. For example in spectroscopy, where data are curves measured as functions of wavelength, functional linear regression and ANOVA models with B-splines expansions have been used in Saeys, De Ketelaere, and Dairus (2008). Kayano, Dozono, and Konishi (2010) use functional cluster analysis to model the three-dimensional (3D) protein structural data that determine the 3D arrangement of amino acids in individual protein. Matsui, Araki, and Konishi (2011) use functional discriminant analysis to classify handwritten characters written in the air with one finger. A successful application of FPCA with Fourier basis expansions has been developed in Valderrama, Ocaña, Aguilera, and Ocaña-Peinado (2010) to forecast airborne cypress pollen concentration from daily evolution of temperatures. Moreover generalizations of classical multivariate techniques to FDA field have been developed as was the case of the formulation and estimation of functional PLS regression from basis expansion of sample curves boarded in Aguilera, Escabias, Preda, and Saporta (2010).

In this paper we consider the related problem of prediction of a categorical response representing the class membership associated to an observed curve. This problem could be also seen as a problem of curve classification or discrimination. This has been the objective of many researchers in recent years from different points of view. Nonparametric FDA methodologies for curve classification of spectrometric food data were used by Ferraty and Vieu (2003). FDA was also used by Preda, Saporta, and Lévéder (2007) from the classical point of view of functional regression methods based on PLS to make a classification of the quality of cookies from their resistance of dough. Linear discriminant analysis has also been generalized for functional data classification (James and Hastie 2001). A hidden process regression model for functional data was used by Chamroukhi, Samé, Govaert, and Aknin (2010) based on an experimental study for curve discrimination. As alternative to the classical multivariate methods of classification are the regression methods for binary response (logistic regression model) or multi-category response (multi-logit regression models) as can be seen in Hervás, Silva, Gutiérrez, and Serrano (2008). These models have the advantage that more than a pure classification method they also allow modeling variables from a set of predictors and the interpretation of the relationship between the categorical response and the functional predictor via the parameters of the model.

The logit regression model for predicting a binary response has been generalized to the case of a functional response variable (Escabias, Aguilera, and Valderrama 2004). Examples of applications of this model in the literature were developed in Ratcliffe, Leader, and Heller (2002) who used it for predicting human foetal heart rate responses to repeated vibroacoustic stimulation, Escabias, Aguilera, and Valderrama (2005) who established the relationship between the risk of drought and time evolution of temperatures, Aguilera, Escabias, and Valderrama (2008) where the probability of lupus flare is estimated from time evolution of stress level and Escabias, Valderrama, Aguilera, Santofimia, and Aguilera-Morillo (2013) where the occurrence of airborne olive pollen peaks is modeled from curves of climatic variables.

The natural generalization of the functional logit model for the case of a categorical response variable with a finite set of categories greater than two is the functional multinomial response model. This model allows curves classification in more than two groups. Different types of logit transformations can be used depending of the type of response (nominal or ordinal). A functional nominal logit model has been considered for predicting land use with the temporal evolution of coarse resolution remote sensing data (Cardot, Faivre, and Goulard 2003). In that paper authors propose a quadrature method to approximate the linear predictor of the model from discrete data and functional PCA to reduce the dimension of the problem, expressing the the functional parameters in terms of spline interpolated eigenfunctions. In this paper we propose a different approach for base-line category logit models based on basis expansions of the functional predictor and the functional parameters that turns the functional model into a multiple one. This consideration usually provides models with highly prediction ability but non-smooth functional parameters. Two different functional PCAs of the functional predictor will be considered to improve the estimation of the functional parameters in the sense of being smooth. Different methods of selection of components will also be introduced, which take into account not only their explained variability but also their ability to predict the response and to provide the best estimation of the functional parameters.

Three sections will be used to explain the methodology proposed in the first one. The second section presents the model formulation and its estimation based on basis expansion methods. The third section contains an alternative estimation based on different types of functional principal component analysis. This method allows to reduce the dimension of the model and to solve multicollinearity problems that arise due to the use of basis expansion methods. The proposed PCA approaches and the model selection procedures will be compared on simulated and real data in the fourth, fifth and sixth sections.

2. Theoretical Framework on Functional Nominal Response Models

In order to formulate the multinomial response model let us consider a sample of observations $\{x_i(t) : t \in T, i = 1, \ldots, n\}$ of a functional predictor $\{X(t) : t \in T\}$, whose sample curves belong to the space $L^2(T)$ of square integrable functions on T.

Let $\{(y_{i1}, \ldots, y_{iS})' : i = 1, \ldots, n\}$ be a set of n vectors sampled from a categorical response variable Y with S categories, defined for each $s = 1, \ldots, S$ by

$$y_{is} = \begin{cases} 1 & \text{if category } s \text{ is observed for } X(t) = x_i(t) \\ 0 & \text{other case} \end{cases}$$

so that each observation is generated by a multinomial distribution $M(1; \pi_{i1}, \ldots, \pi_{iS})$ with $\pi_{is} = P[Y = s | X(t) = x_i(t)]$ and $\sum_{s=1}^{S} \pi_{is} = 1 \quad \forall i = 1, \ldots, n$.

As in any multinomial regression model (see Agresti 2002), y_{iS} is redundant, so let denote by $y_i = (y_{i1}, \ldots, y_{i,S-1})'$ the response vector for subject i, with mean vector $\mu_i = E[Y_i] = (\pi_{i1}, \ldots, \pi_{i,S-1})'$.

The base-line category logit model for nominal responses can be extended to the functional case as a functional generalized linear model (James 2002) whose link function is given by the baseline-category logit transformations l_{is} that pairs each response with a baseline category (usually the last one)

$$l_{is} = \log [\pi_{is}/\pi_{iS}] = \alpha_s + \int_T x_i(t) \beta_s(t) \, dt.$$

Then, the probabilities of the multinomial response are modeled in terms of the functional predictor and the parameters (functional and non-functional) as

$$\pi_{is} = \frac{\exp\left\{\alpha_s + \int_T x_i(t) \beta_s(t) \, dt\right\}}{\sum_{s=1}^{S} \exp\left\{\alpha_s + \int_T x_i(t) \beta_s(t) \, dt\right\}}, \quad s = 1, \ldots, S, \ i = 1, \ldots, n, \tag{1}$$

with $\alpha_S = 0 = \beta_S(t)$.

This model formulation leads to an interpretation of the relationship between the nominal response and the functional predictor. For example, the exponential of the integral $\int_{t_0}^{t_0+h} \beta_s(t) g(t) \, dt$ is the multiplicative change in the odds of response $(Y = s)$ against response $(Y = S)$ provided by a change of the curve $x_i(t)$ according to a function $g(t)$ in the interval $[t_0, t_0 + h]$ (see Aguilera, Escabias and Valderrama (2008) for a detailed explanation of this type of interpretation with the binary response logit model).

The estimation of the parameters of this model is an ill-posed problem due to the infinite dimension of the predictor space (see Ramsay and Silverman (2005) for a discussion on the functional linear model). In addition, the functional predictor can not be observed continuously in time but in a set of discrete time points $\{t_{ik} : k = 1, \ldots, m_i\}$ that could be different for each sampled individual. Basis expansion methods allow us to overcome these two problems. The most used approach is to reduce dimension by performing an orthonormal basis expansion of the functional predictor. Then, a finite number of coefficients of such orthonormal representations are used as predictor variables of the functional regression model that becomes a multiple regression model whose parameters are estimated as usual by least squares or maximum likelihood. This dimension reduction approach has been studied by Müller and Stadtmüller (2005) in the theoretical framework of functional generalized linear models where asymptotic tests and simultaneous confidence bands for the parameter function have been obtained. On the other hand, an alternative estimation procedure based on B-spline expansion of the functional parameter (but not the predictor curves) that maximizes the penalized log-likelihood has been studied in Marx and Eilers (1999) for a functional binomial response model and in Cardot and Sarda (2005) for the general context of functional generalized linear models.

Let us consider that both the predictor curves and functional parameters belong to a finite space generated by a basis of functions

$$x_i(t) = a_i'\Phi(t), \ \beta_s(t) = \beta_s'\Phi(t),$$

with $\Phi(t) = (\phi_1(t), \ldots, \phi_p(t))'$ being a vector of basic functions that generate the space where $x(t)$ belongs to, and $a_i = (a_{i1}, \ldots, a_{ip})'$ and $\beta_s = (\beta_{s1}, \ldots, \beta_{sp})'$ being the vectors of basis coefficients of the sample curves and parameter functions, respectively.

This way the functional model turns to a multiple one given by

$$l_{is} = \alpha_s + \int_T x_i(t)\beta_s(t)\,dt = \alpha_s + a_i'\Psi\beta_s$$
$$s = 1, \ldots, S-1, \quad i = 1, \ldots, n,$$

with $\Psi = (\psi_{uv})$ being the $p \times p$ matrix of inner products

$$\psi_{uv} = \int_T \phi_u(t)\phi_v(t)\,dt.$$

In matrix form each vector of logit transformations $L_s = (l_{1s}, \ldots, l_{ns})'$ can be expressed as

$$L_s = \alpha_s\mathbf{1} + A\Psi\beta_s, \ s = 1, \ldots, S-1. \tag{2}$$

In spite of having considered the same type of basis for sampled curves and functional parameters, they could be different. In this case Ψ matrix would have as entries the between-basis inner products.

The estimation of this multicategory logit model is carried out by maximizing the multinomial log likelihood

$$\log\left[\prod_{i=1}^{n}\left[\prod_{s=1}^{S}\pi_{ij}^{y_{is}}\right]\right]$$

under model (1). Newton-Raphson is the most used method that yields the ML parameter estimates because of the concavity of the log-likelihood equation.

Before estimating the parameters of this multiple model (basis coefficients of the parameter functions) it is mandatory to obtain the sample curves basis coefficients from the sample information $x_i = (x_{i1}, \ldots, x_{im_i})$ available for each sample curve $x_i(t)$ at a discrete set of knots $(t_{i1}, \ldots, t_{im_i})$. Different numerical procedures as interpolation (data observed without error) or least squares approximation (noisy data) can be used to compute these basis coefficients in practice. Natural cubic spline interpolation was first considered to estimate functional PCA (Aguilera, Gutiérrez, and Valderrama 1996). Quasi natural cubic spline interpolation has been introduced to estimate the risk of drought from time evolution of temperatures (Escabias, Aguilera, and Valderrama 2005). Least squares approximation with both B-splines and trigonometric functions has been recently used for interpreting the relationship between time evolution of stress and flares in Systemic Lupus Erythematosus patients (Aguilera, Escabias, and Valderrama 2008).

3. Functional PCA Based Solutions

As set out in the two previous sections, in addition to a classification method, multinomial logit models allow to interpret the relationship between the categorical response variable (that defines the groups where the predictor variable observations belong to) and the functional predictor. This interpretation is carried out by using the slope parameters of the model, that in the case of functional multinomial logit model are a set of functions. With this objective in mind an accurate and interpretable estimation of these functional parameters is very important with interpretability being synonymous of smoothness in functional data analysis. As in the binary response case (Escabias, Aguilera, and Valderrama 2004), basis expansion estimation of logit models usually provides good predictions of the response and consequently a fair classification rule but an inaccurate parameter estimation with high variability (due to multicollinearity). It makes the interpretability difficult of the relationship between variables so that the equilibrium between

smooth estimation of the (functional) parameters of the model and accurate prediction of the response is our aim in this paper.

A well-known solution to reduce dimension and solve multicollinearity problems in functional data analysis is to use as regressors a reduced set of uncorrelated components instead of the columns of the $A\Psi$ matrix. In the particular case of binary response, we can find in the literature different approaches based on PCR and PLS (Escabias, Aguilera, and Valderrama 2007). Two different functional PCA approaches that agree with orthonormal basis functions are compared in this paper to get accurate estimations of both the functional parameter and the categorical nominal response.

3.1   Overview on Functional PCA

In this section we will present a brief summary of basic ideas on functional PCA obtained as a generalization of multivariate PCA to the functional case (data are curves). Without loss of generality and in order to clarify the theoretical aspects, we consider in this summary that curves are centered, that is, $\overline{x}(t) = \sum_{i=1}^{n} x_i(t) = 0$.

Given a sample of centered curves $\{x_i(t) \in L^2[T] : i = 1, \ldots, n\}$, the j-th principal component is given by

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

where the weight function f_j is obtained by maximizing

$$Max_f Var\left[\int_T x_i(t) f(t)\, dt\right]$$

$$r.t. \left\{\|f\|^2 = 1 \ and \ \int f_\ell(t) f(t)\, dt = 0 \quad \forall \ell = 1, \ldots, j-1\right\}.$$

The weight functions are obtained as the eigenfunctions of the covariance operator C defined by

$$Cf(s) = \int c(t, s) f(t)\, dt \quad s \in T$$

in terms of the sample covariance function

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

That is, $Cf_i = \lambda_i f_i$. The principal components ξ_i are uncorrelated and their variances are given by the eigenvalues $Var[\xi_i] = \lambda_i$.

Then, the sample curves admit the following principal component decomposition

$$x_i(t) = \sum_{i=1}^{n-1} \xi_{ij} f_j(t). \tag{3}$$

By truncating this representation in terms of the first q principal components, we can obtain an approximation of the sample curves whose explained variance is given by $\sum_{i=1}^{q} \lambda_i$.

Let us now suppose that the sample paths belong to a finite dimension space spanned by a basis $x_i(t) = a_i'\Phi(t)$ with $\Phi(t) = (\phi_1(t), \ldots, \phi_p(t))'$ being the vector of basic functions. Then, the theoretical results shown in Ocaña, Aguilera, and Escabias (2007) set up the following equivalences between different functional and multivariate PCAs with respect to the usual inner products in $L^2[T]$ and $\mathbb{R}^p$.

1. Multivariate PCA of $A\Psi$ matrix with respect to the usual metric in $\mathbb{R}^p$ is equivalent to functional PCA of the transformed sample curves $L(x_i) = \Phi'\Psi^{1/2}a_i$ with respect to the usual metric in $L^2[T]$.

   The matrix whose columns are the principal components is given by $\Gamma = A\Psi V$ with V being the matrix of eigenvectors ($VV' = I$) of the sample covariance matrix of $A\Psi$. Then, the eigenfunctions of the principal component decomposition given by equation (3) admits a basis expansion given by $f_j = \Phi'\Psi^{-1}v_j$, with v_j being the columns of the matrix V of eigenvectors.

2. Functional PCA of the sample curves with respect to the usual inner product in $L^2(T)$ is equivalent to multivariate PCA of matrix $A\Psi^{1/2}$ with respect to the usual inner product in $\mathbb{R}^p$.

   In this case the matrix of principal components is given by $\Gamma = A\Psi V = A\Psi^{1/2}U$ where $V = \Psi^{-1/2}U$ and U is the matrix of eigenvectors ($UU' = I$) of the sample covariance matrix of $A\Psi^{1/2}$. The basis expansion expression for the eigenfunctions is now given by $f_j = \Phi'\Psi^{-1/2}u_j$, where u_j are the columns of the eigenvector matrix U.

Let us observe that these two versions of functional PCA match when the basis Φ is orthonormal. In this case, functional PCA with the usual metric in $L^2[T]$ is equivalent to multivariate PCA of the matrix A of sample curves basis coefficients with respect to the usual metric in $\mathbb{R}^p$.

Different basis (B-splines, wavelets, trigonometric, ...) can be used depending of smoothness and performance of sample functions.

3.2 Functional Principal Component Estimation

Following the idea of principal component regression (Massy 1965), in this paper we propose to use as regressors of the base-line category response model, a reduced set of functional principal components of the sample curves. We will consider the two different functional PCAs described above that will be referred as PCA1 and PCA2 based solutions, respectively.

Let $\Gamma = A\Psi V$ be a matrix of functional principal components computed by using PCA1 or PCA2 approaches. Then, the multinomial logit model (2) can be equivalently expressed in terms of all principal components as

$$L_s = \alpha_s \mathbf{1} + \Gamma\gamma_s, , \tag{4}$$

so that the ML estimation of the basis coefficients of the functional parameters can be obtained through the estimation of the parameter of model (4) by

$$\widehat{\beta}_s = V\widehat{\gamma}_s.$$

In spite of the orthogonality of functional principal components, the exact equivalence between models 2 and 4 makes the estimations to be identical and the problems of multicollinearity to be kept. In the simulation study developed in the next section, we will see that the ML estimation of the functional parameter provided by using all PCs as predictor variables is very rough and inaccurate. Because of this we propose to use as predictors an optimum set of m PCs contained in the columns of matrix $\Gamma^{(m)} = (A\Psi)V^{(m)}$. Then, the functional principal component nominal response logit model is given by

$$L_s^{(m)} = \alpha_s^{(m)}\mathbf{1} + \Gamma^{(m)}\gamma_s^{(m)},$$

and provides a ML estimation of the functional parameter given by

$$\hat{\beta}_s^{(m)}(t) = \Phi'(t)\widehat{\beta}_s^{(m)}$$

with $\widehat{\beta}_s^{(m)} = V^{(m)}\widehat{\gamma}_s^{(m)}$.

3.3 Principal Components Selection

An important problem in principal component regression is to select the optimum number of functional principal component (of any type) to use and the order in which they must be included in the model. Following the methodology considered in Aguilera, Escabias, and Valderrama (2006) for the multiple binary logit model, in this paper we have considered with each type of PCA two different ways for including PCs in the model. The first one consists of including principal components in the order given by their

explained variability. In the second one, PCs are included by a forward stepwise method based on conditional likelihood ratio test that takes into account their relationship with the response variable.

In relation to the number of components to retain two methods will be compared. the first one is the classical method used in principal component regression based on minimization the leave-one-out prediction error or the leave-one-out misclassification rate via cross-validation. The second consist of selecting the number of components that provide the most accurate estimation of the functional parameters. Different measures can be used to evaluate the accuracy of estimated functional parameters and the fit of the different considered models.

As goodness of fit measures we have considered

- The correct classification rate (CCR) defined as the rate of agreements between the observed and predicted category of the response, by considering the predicted category of an individual the one associated with the highest predicted probability.

- The probabilities mean square error ($PMSE$) for simulated examples defined as

$$PMSE(m) = \frac{1}{S}\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n}\left(\pi_{is} - \widehat{\pi}_{is}^{(m)}\right)^2.$$

- The leave-one-out cross-validation mean squared error of prediction defined as

$$CVMSE(m) = \frac{1}{S}\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n}\left(y_{is} - \widehat{\pi}_{(-i)s}^{(m)}\right)^2,$$

where $\widehat{\pi}_{(-i)s}$ is the probability of the s category predicted for the ith individual by using the model fitted after removing the ith individual from the data.

- The leave-one-out cross-validation correct classification rate ($CVCCR$) defined as the rate of agreements between the observed category for an individual and the predicted category (that associated with the highest predicted probability) with the model estimated without taking into account that individual in the estimation process.

As accuracy measures of the estimation of the functional parameters we have considered

- The mean of the integrated mean square error of the parameter functions ($IMSE$) for simulated examples defined as

$$IMSE(m) = \frac{1}{S-1} \sum_{s=1}^{S-1} \frac{1}{T} \int_T \left(\beta_s(t) - \widehat{\beta}_s^{(m)}(t) \right)^2 dt.$$

- The estimated variance of the estimated parameter functions (VAR) given by the mean of the sum of variances of intercepts and parameter function basis coefficients

$$VAR(m) = \frac{1}{S-1} \sum_{s=1}^{S-1} \left[Var[\widehat{\alpha}_s^{(m)}] + \sum_{j=1}^{p} Var[\widehat{\beta}_{sj}^{(m)}] \right],$$

where $Var[\widehat{\alpha}_s^{(m)}]$ and $Var[\widehat{\beta}_{sj}^{(m)}]$ are the variance of the estimators, obtained in the estimation process as a sub-product of the Newton-Raphson method. Their estimates should be used in practice.

The optimum number of principal components can be fixed by two different types of criteria; the response prediction-type criterium and the functional parameters-type criterium. Taking into account the response, the classic cross-validation method selects the number of components that optimizes (maximize or minimize) some leave-one-out error term (*CVMSE* or *CVCCR*). From this point of view we can consider as optimum the number of components that firstly locally minimize the *CVMSE* or maximize the *CVCCR*. On the other hand, thinking in the functional parameter-type criterium we should consider as optimum the number of components that minimize the *IMSE* because it provides the best estimation of the functional parameters. Associated to the estimated parameters are their estimated variances (*VAR*), that can help us to detect excessive fluctuations that denote inaccuracy.

4.   Simulation Study

In this section, we will present the results of a simulation study developed to show how functional principal component analysis improves the estimation of the functional parameters of a functional multinomial logit model with nominal response. The two types of functional PCA previously considered and the two methods of inclusion of components in the regression model are compared. The ability of leave-one-out cross-validation for selecting the number of PCs that provides the most accurate estimation of the functional parameters is also discussed.

The functional predictor considered in the study is the stochastic process

$$X(t) = Z(t) + \frac{t}{4} + 5B,$$

All curves

$\beta_1(t)$ & $\widehat{\beta}_1(t)$

$\beta_2(t)$ & $\widehat{\beta}_2(t)$
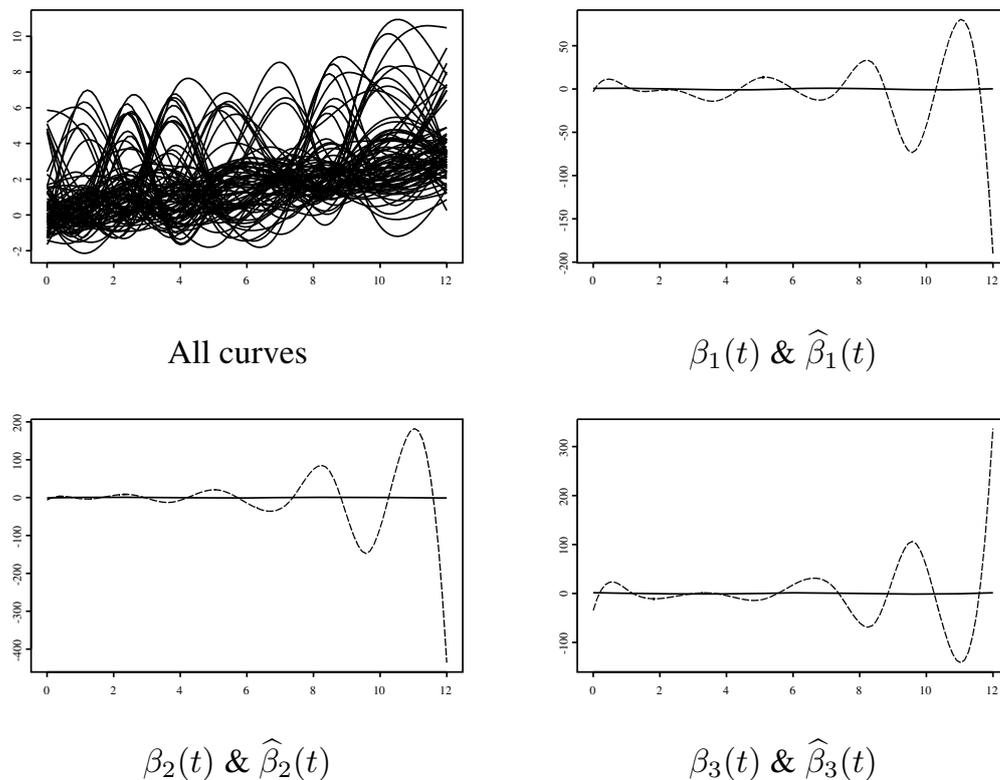
$\beta_3(t)$ & $\widehat{\beta}_3(t)$

Figure 1. Simulated curves of the functional predictor variable (top left). Original trigonometric parameter functions (solid line) and their basis expansion estimation without using PCA (dashed line) for the functional multinomial logit model

with $Z(t)$ being a centered gaussian process that has as covariance function $C(s,t) = 0.5^{80|t-s|}$ and B being a Bernouilli random variable with probability 0.1. The domain of the considered functional variable is $T = [0,12]$ interval.

Taking into account that usually it is impossible to record the functional form of the curves of a functional variable and we only have discrete observations at different points of the domain interval, $n = 80$ curves have been simulated from the defined process at the set of unequally-spaced nodes $\Pi = \{0, 1.1, 2.5, 3.7, 5.1, 7.3, 8.5, 9.6, 12\}$. The true functional forms of the curves have been reconstructed via quasi-natural cubic spline interpolation and they can be seen in Figure 1.

A nominal random variable with four categories was considered as the response variable. The response probabilities have been simulated by using the functional multinomial logit model for nominal response (1) in terms of cubic B-spline expansion of sample curves and parameter functions. The intercept parameter is $(\alpha_1, \alpha_2, \alpha_3) = (0.30, 0.19, 0.20)$ and the slope functional parameters are the quasi-natural cubic spline interpolation at the knots of partition Π of sinusoidal functions

$$\beta_1\left(t\right) = \cos\left(t - \frac{\pi}{4}\right)$$
$$\beta_2\left(t\right) = \sin\left(t - \frac{\pi}{4}\right)$$
$$\beta_3\left(t\right) = \cos\left(t - \frac{\pi}{4}\right) - \sin\left(t - \frac{\pi}{4}\right).$$

Finally, the observations of the response are randomly simulated from a multinomial distribution with the simulated probabilities as parameters.

After data simulation the nominal response functional logit model (2) was fitted, the parameter functions reconstructed (see Figure 1) and the goodness of fit statistics as the correct classification rate (CCR) and *PMSE* calculated. All these measures suggested a good prediction ability of the model (*CCR*=87.5, *PMSE*=0.0471) but the form of the estimated functional parameters and the accuracy measure (*IMSE*=3.94E+4) showed a bad estimation of the functional parameters. The problem of multicollinearity is clear in this example with estimated variance of the estimated parameter functions very high (*VAR*=801.979).

In order to improve the estimation of the functional parameters two functional principal component solutions (PCA1 and PCA2) were adopted. The nominal response models with different number of functional PCs of each type as predictors were fitted by using two orders for including PCs in the model (variability and forward stepwise). Then, the parameter functions basis coefficients were approximated in each case and the accuracy measures calculated. The results can be seen in Tables 1 and 2 for PCA1 and PCA2, respectively. Figures 2, 3 and 4 display the graphs of the estimation of the functional parameters en terms of different number of PCs included in the model with variability and stepwise orders.

To select the optimal model (number of PCs) in each case two main criteria were considered. In the first the number of PCs is select by minimizing the *IMSE* and in the second by the first local minimum of the *CVMSE*. Let us observe that in the case of PCA1 the number of components selected with these two criteria with both orders of inclusion of components in the model (variability and stepwise) is very similar. However, in the case of PCA2, the number of components selected by minimizing *CVMSE* is lower and, as a consequence, the *PMSE* and *IMSE* errors are bigger. Let us also observe that the best parameter function estimation provides also the best prediction of probabilities with both PCA approaches. The criterium based on the first local maximum of the *CVCCR* is not discussed because it gives results very similar to *CVMSE*.

Another important aspect observed in both tables is that the best estimation of the parameter function is sometimes followed by a big increment in the variance of the estimated parameters. Because of this we have also

Table 1. Goodness of fit measures (*CCR* and *PMSE*) and accuracy measures (*IMSE*) of PCA1-type models in terms of different number of FPCs included by variability and stepwise orders. Each row presents the measures associated to the model with as many FPCs as indicates the first column. The cumulated variance of the different number of FPCs is also included. The last two columns present the results of cross-validation.

			Variability order					
s		*CumVar*	*IMSE*	*Var*	*RIVAR*	*CCR*	*PMSE*	*CVCCR*	*CVMSE*
1		24.79	6.9735	3.5614	0.82	61.25	0.0954	61.25	0.1502
2		43.12	7.0011	6.4650	0.15	66.25	0.0870	66.25	0.1372
3		57.53	6.6084	7.4094	0.39	70.00	0.0747	67.50	0.1269
4		71.23	5.6383	10.3300	1.67	71.25	0.0624	66.25	0.1208
5		83.19	1.2354	27.6207	0.63	76.25	0.0410	63.75	0.1180
6		90.61	2.6581	44.8871	0.21	80.00	0.0400	63.75	0.1150
7		95.23	2.7261	54.3180	0.83	77.50	0.0408	65.00	0.1237
8		98.19	6.0630	99.3756	0.94	78.75	0.0385	63.75	0.1360
9		100.0	16.6856	192.5328	0.00	82.50	0.0399	63.75	0.1398
10		100.0	19.3837	192.8865	3.16	82.50	0.0399	66.25	0.1424
11		100.0	3.94E+4	801.9790		87.50	0.0471	66.25	0.1387

			Stepwise order					
s	pc's	*CumVar*	*IMSE*	*Var*	*RIVAR*	*CCR*	*PMSE*	*CVCCR*	*CVMSE*
1	2	18.3310	7.2975	3.8663	0.21	67.50	0.0904	66.25	0.1365
2	5	30.2940	6.5001	4.6843	0.43	66.25	0.0725	63.75	0.1250
3	3	44.7006	5.3616	6.7150	0.44	67.50	0.0583	68.75	0.1169
4	4	58.3972	4.1622	9.7008	0.09	73.75	0.0494	66.25	0.1107
5	10	58.3979	4.40E+4	10.5996	0.21	78.75	0.0549	63.75	0.1180
6	6	65.8190	2.47E+4	12.8172	0.94	81.25	0.0519	63.75	0.1150
7	8	68.7849	5.01E+4	24.8142		81.25	0.0534	61.25	0.1270

considered an "ad hoc" selection model criterion based on quantifying the relative importance of this increment. So, we will select as optimal the number s of PCs such that $RIVAR(s) = [VAR(s+1) - VAR(s)]/VAR(s)] \geq 1$ what would mean a doubled increment in the variance.

In order to extract conclusions on the good performance of the FPCA-based estimation approaches, we have repeated this simulation 500 times: simulation of predictor curves and response, computation of principal components with PCA1 and PCA2 approaches, fit of multinomial models with different number of principal components and different inclusion orders (variability and stepwise), reconstruction of the functional parameters's basis coefficients and computation of the different accuracy measures. For each repetition and in each one of the four kinds of fit (Variability order with PCA1, stepwise order with PCA1, Variability order with PCA2 and stepwise order with PCA2) we have considered three optimal models (number of principal components) according to the minimum *IMSE*, the first min-

Table 2. Goodness of fit measures (*CCR* and *PMSE*) and accuracy measures (*IMSE*) of PCA2-type models in terms of different number of FPCs included by variability and stepwise orders. Each row presents the measures associated to the model with as many FPCs as indicates the first column. The cumulated variance of the different number of FPCs is also included. The last two columns present the results of cross-validation.

				Variability order				
s		*CumVar*	*IMSE*	*Var*	*RIVAR*	*CCR*	*PMSE*	*CVCCR*	*CVMSE*
1		18.57	6.8012	2.6425	0.85	61.25	0.0946	60.00	0.1491
2		33.07	5.1100	4.8888	0.29	68.75	0.0578	58.75	0.1216
3		46.26	4.7895	6.2946	1.82	70.00	0.0552	66.25	0.1149
4		58.21	4.7710	17.7397	0.28	75.00	0.0457	68.75	0.1057
5		69.24	4.0829	22.6267	0.89	73.75	0.0447	66.25	0.1093
6		79.50	2.0182	42.6916	0.62	77.50	0.0376	65.00	0.1227
7		88.59	4.4536	69.2549	0.35	76.25	0.0382	63.75	0.1275
8		95.34	3.6707	93.1771	2.33	77.50	0.0379	63.75	0.1393
9		99.98	16.6261	310.7311	3.3E+6	82.50	0.0399	63.75	0.1398
10		100.0	3.24E+5	1.04E+9	0.61	87.50	0.0498	68.75	0.1413
11		100.0	4.26E+5	1.67E+9		90.00	0.0509	67.50	0.1432
				Stepwise order				
s	*pc's*	*CumVar*	*IMSE*	*Var*	*RIVAR*	*CCR*	*PMSE*	*CVCCR*	*CVMSE*
1	2	14.51	5.6082	2.9154	2.56	62.50	0.0609	57.50	0.1190
2	4	26.45	6.0207	10.3871	0.07	71.25	0.0524	65.00	0.1102
3	3	39.64	5.6558	11.1028	0.31	73.75	0.0504	67.50	0.1031
4	6	49.90	4.5572	14.5873	6.6E+6	73.75	0.0475	68.75	0.1057
5	10	49.92	2.48E+5	9.57E+7	11.85	72.50	0.0517	62.50	0.1195
6	11	49.93	2.22E+5	1.23E+9	0.04	75.00	0.0508	66.25	0.1203
7	8	56.67	2.82E+5	1.28E+9		78.75	0.0521	66.25	0.1239

imum of *CVMSE* and the first time that *RIVAR* is greater than one. In order to compare the results we have obtained box-plots for the different measures (number of PCs, *IMSE*, *CCR*, *CVMSE*, *VAR*) of the optimal models in each combination of PCA and order of inclusion of components. See Figures 5 and 6 for comparing *IMSE* with *CVMSE* and with *RIVAR*, respectively.

Let us observe that with both PCA approaches the results are very similar. In variability order, the number of PCs selected by minimizing *CVMSE* is significantly lower than minimizing *IMSE*. Therefore, the accuracy of the estimated parameter functions is significantly smaller (bigger *IMSE*) and the prediction errors (*CVMSE*, *PMSE* and *CCR*) are also bigger. On the other hand, forward stepwise order selects a similar number of PCs with both, *IMSE* and *CVMSE* criteria with slightly bigger *IMSE* and similar prediction errors. Let us also observe that the number of pc's needed to minimize *IMSE* with stepwise selection is lower than with variability order but the prediction and *IMSE* errors are not significantly larger.

With respect to the criterium based on the relative increment in the estimated variance of the functional parameters, we can see that when PCs
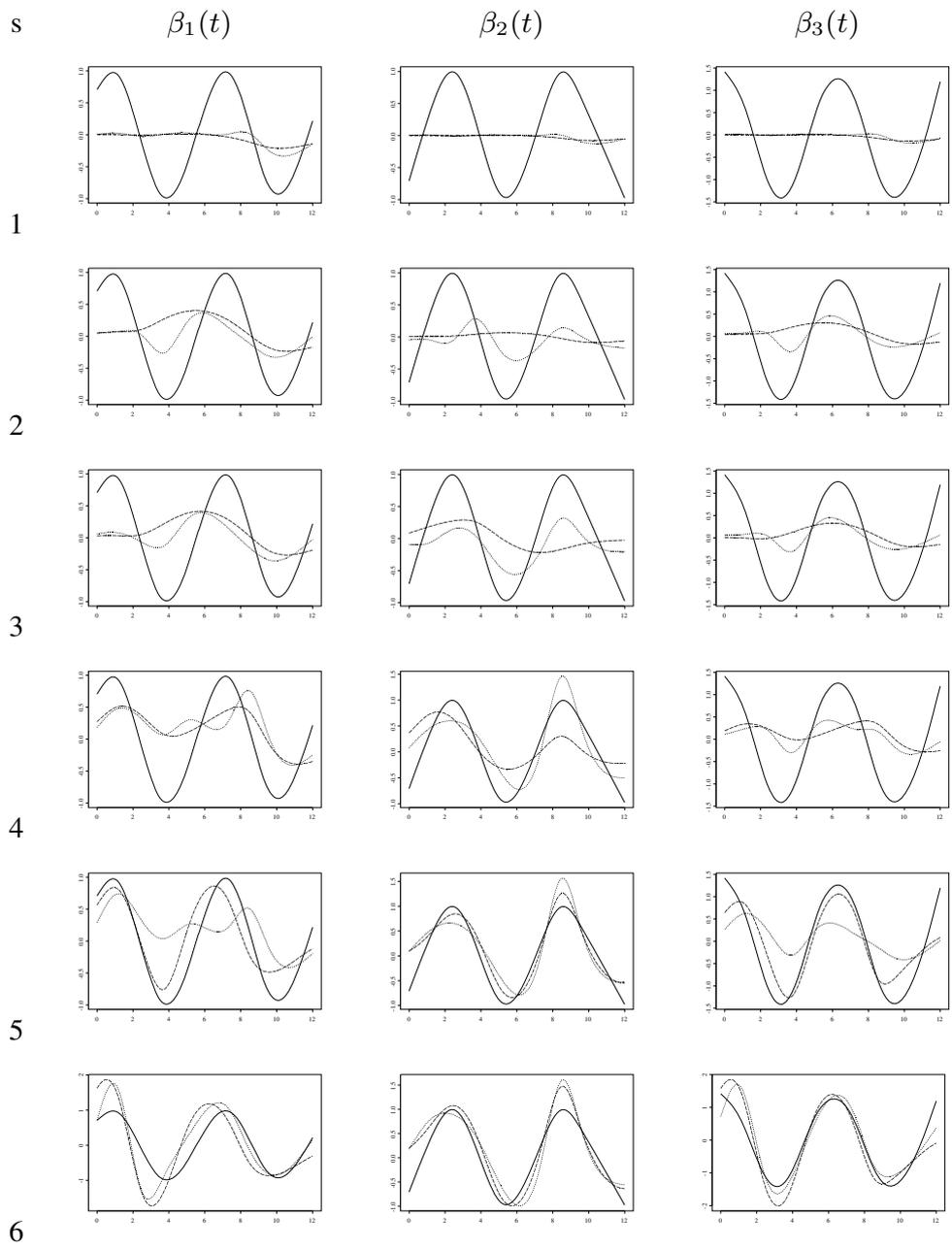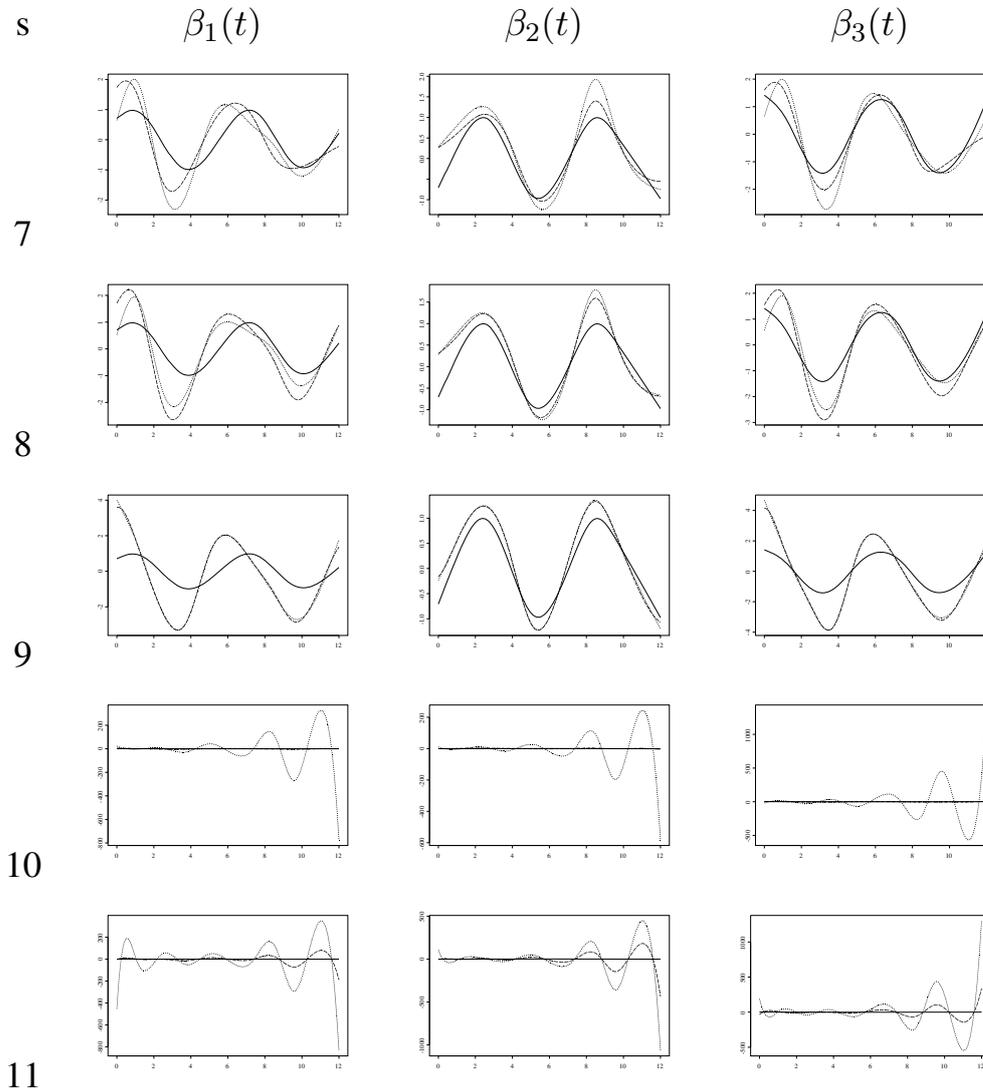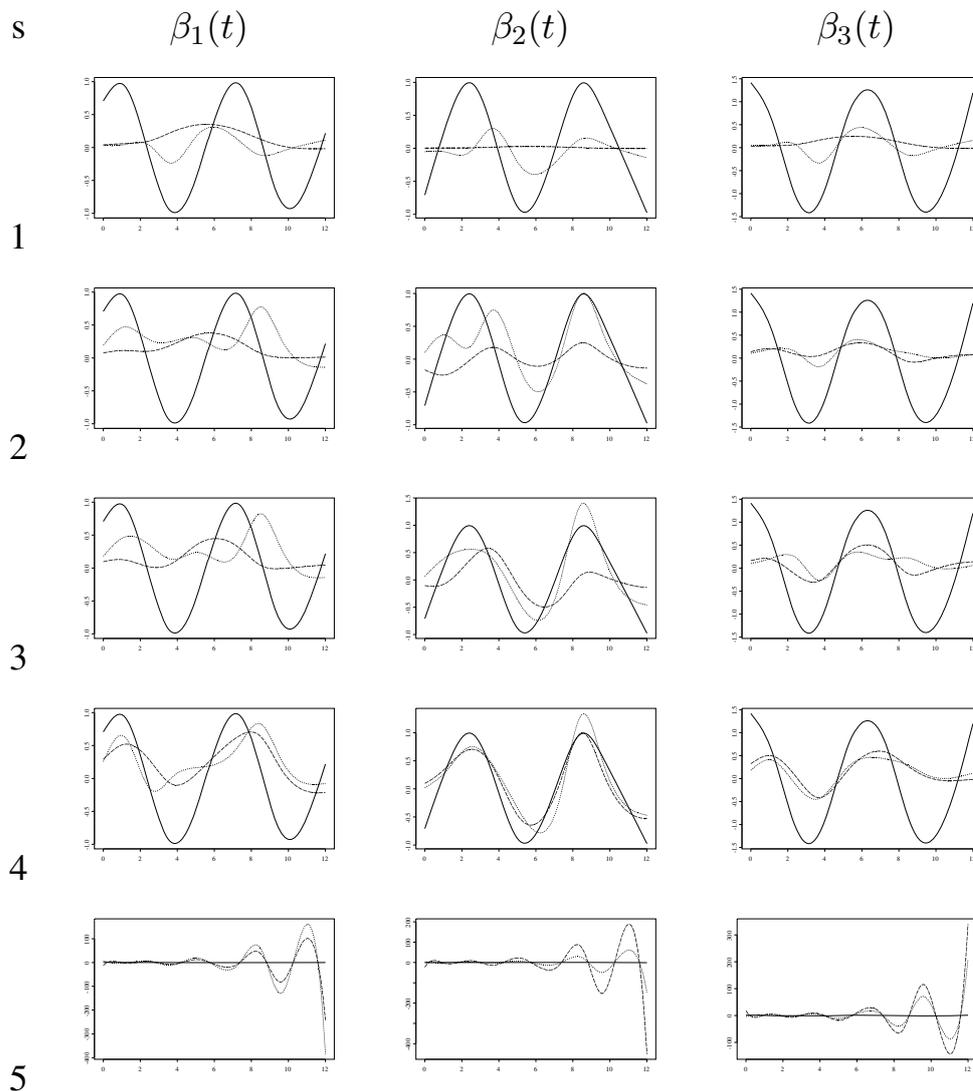
Figure 2. Simulated functional parameters (solid line) and their estimations in terms of different number of principal components included by variability order of PCA1-type (dashed line) and PCA2-type (dotted line).

are introduced in variability order the selected model provides better functional parameter estimation than *CVMSE* criterium with good prediction ability. With stepwise selection the results given by *CVMSE* and *RIVAR* are similar providing in both cases an acceptable degree of accuracy in both parameter estimates and predictions.

Figure 3. Simulated functional parameters (solid line) and their estimations in terms of different number of principal components included by variability order of PCA1-type (dashed line) and PCA2-type (dotted line).

All previous comments are corroborated by the means of the estimated functional parameters in the 500 simulations that are drawn in Figure 7.

In order to compare the classification ability of the functional multinomial logit model with alternative methods as the functional linear discriminant analysis proposed by James and Hastie in 2001, we obtained the correct classification rates provided by this method in the 500 repetitions of the simulation. The fits were obtained thanks to the S-PLUS code that Professor James has free in his personal web page. The classification rates provided by this method are summarized in Table 3. As we can see, our method provides a bit better classification rates.

Figure 4. Simulated functional parameters (solid line) and their estimations in terms of different number of principal components included by stepwise order of PCA1-type (dashed line) and PCA2-type (dotted line).

As conclusion, this simulated example has empirically demonstrated that:

- There is no difference between the type of PCA to improve the estimation of functional parameters of the base-line logit model, the one that is easiest to apply can be chosen.

- The inclusion of functional principal components should be done in the order given by forward stepwise method based on conditional likelihood ratio test instead of variability order, mainly because we get better estimations of the functional parameter with less principal components. And what is more important *IMSE* and *CVMSE* criteria agree to select the optimum estimation of the functional parameter.
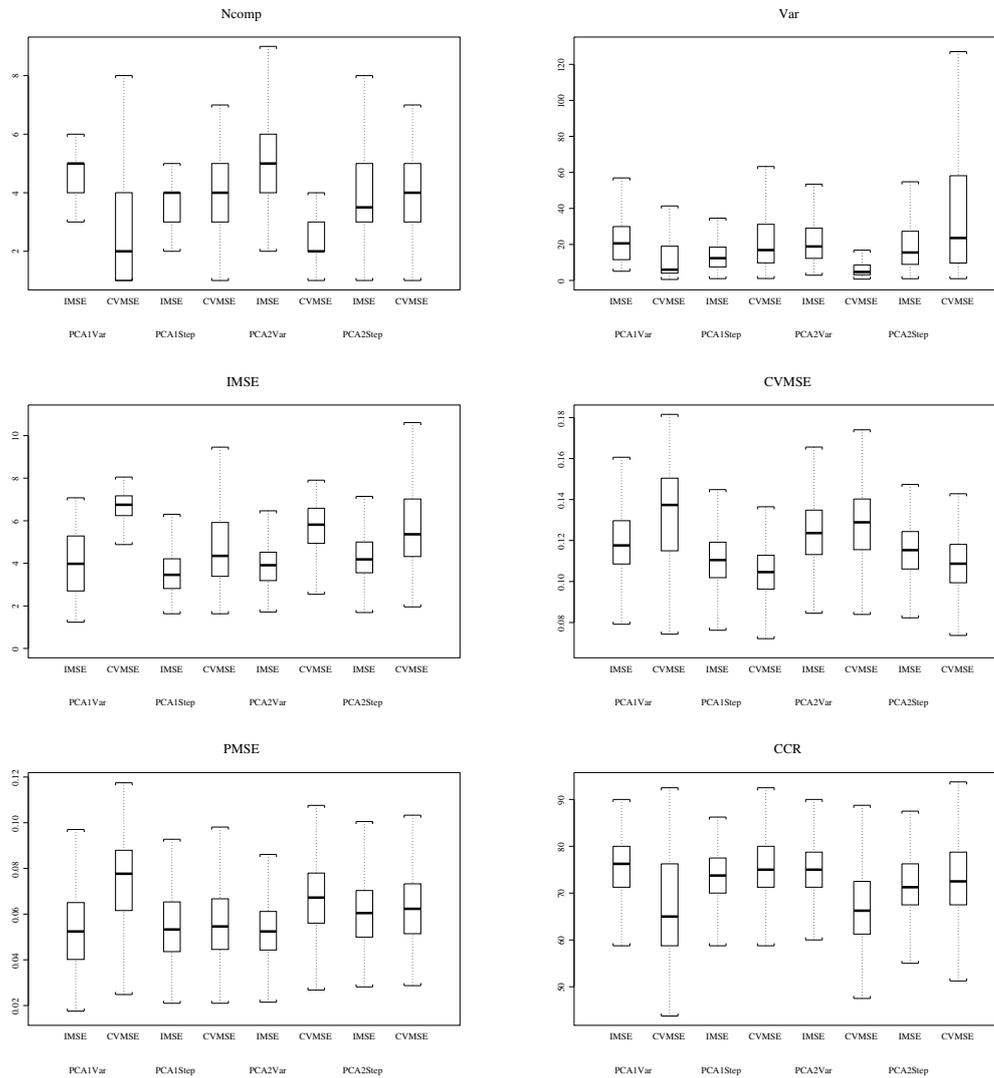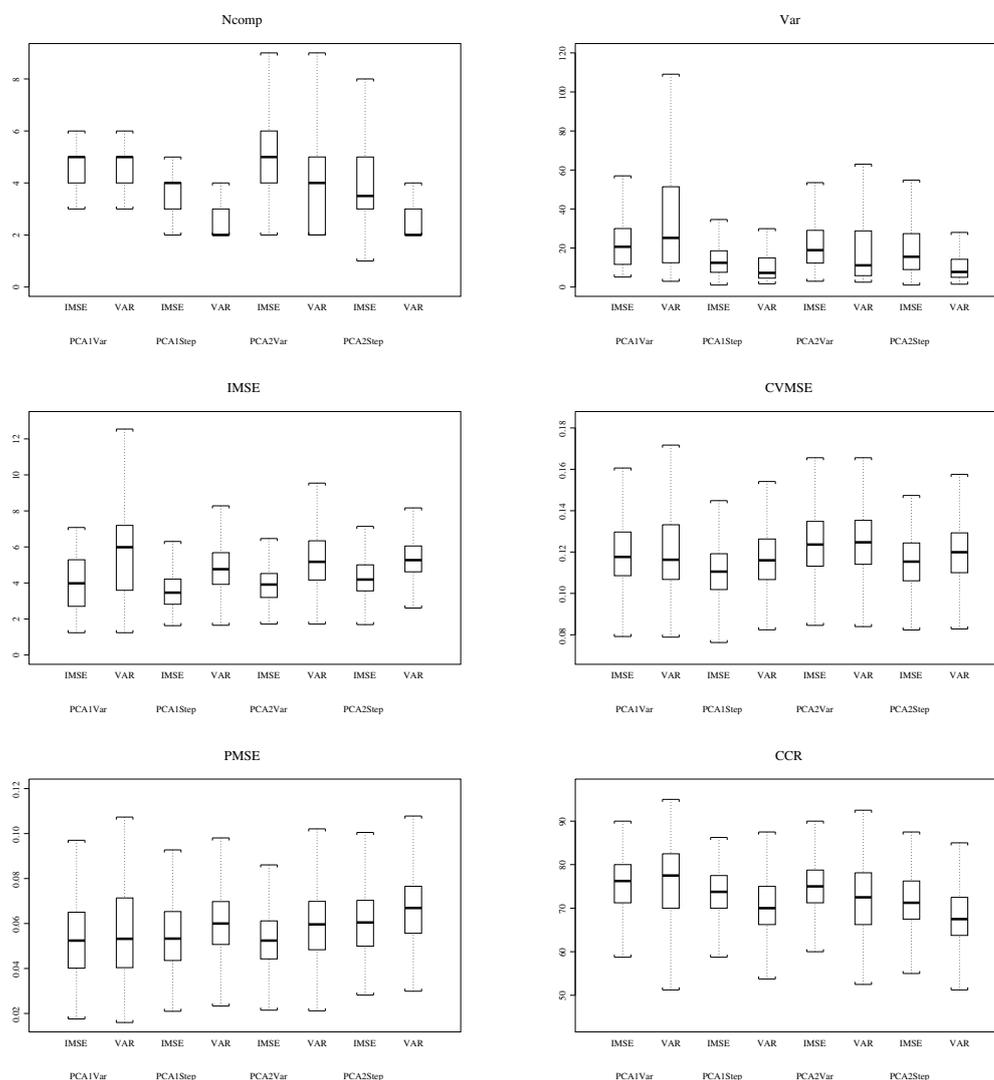
Figure 5. Box-plots of accuracy measures under the two different selection model criteria (*IMSE* and *CVMSE*) with the two different PCA-type (PCA1 and PCA2) and order of inclusion (Var=Variability and Step=Stepwise).

As summary this simulated example has shown that by including principal components in the model in the order given by stepwise method we can use the *CVMSE* and *RIVAR* measures to find an accurate estimation of the functional parameters in real data examples where it is impossible to calculate the *IMSE*.

5. Classification of Spectrometry Data

As was stated in the introduction section, spectrometric data consist of curves of spectrometry (absorbance measured in terms of wavelength) of different substances as for example food. In spectroscopy the most common problem is calibration that consists of estimating an scalar response variable from the spectrum. Despite of the functional nature of spectra data this prob-

Figure 6. Box-plots of accuracy measures under the two different selection model criteria (*IMSE* and *RIVAR*) with the two different PCA-type (PCA1 and PCA2) and order of inclusion (Var=Variability and Step=Stepwise).

lem is usually analyzed with multivariate statistical methods such as Principal Component Regression (PCR) and Partial Least Squares Regression (PLS) that consider the spectrum as a vector associated with its measures into a finite number of wavelengths. Taking into account that the absorbance at two nearby wavelengths are highly correlated, it could provide better estimations the consideration of the spectrum as a curve instead of a vector. The potential use of functional data analysis in spectroscopy and chemometric data was stated by Saeys, De Ketelaere, and Dairus (2008).
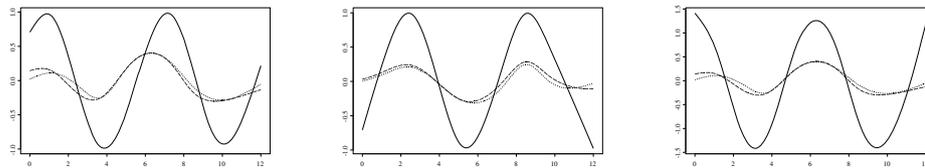
Chemometricians are also interested in the classification of chemometric data (curves of spectrum) according to a characteristic of interest of the substance that generated the curve (see for example Ferraty and Vieu
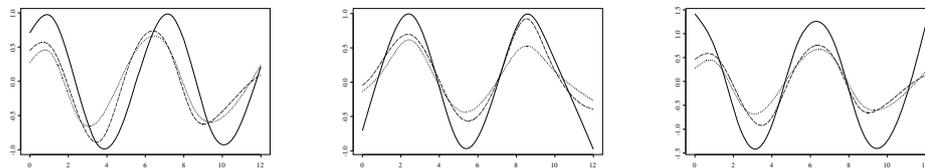
$$\beta_1(t) \qquad \beta_2(t) \qquad \beta_3(t)$$
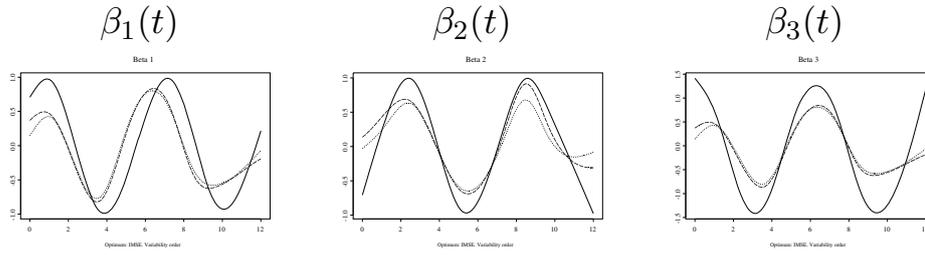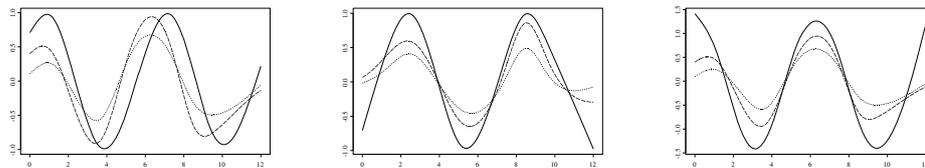
IMSE criterium with variability order
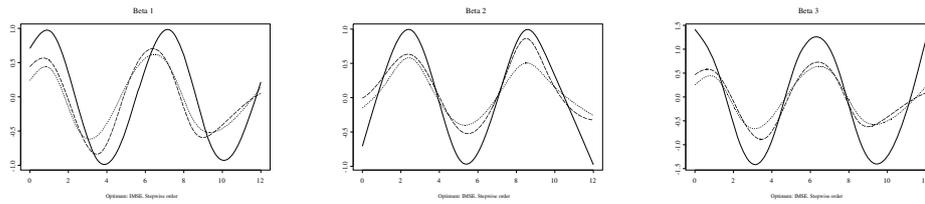
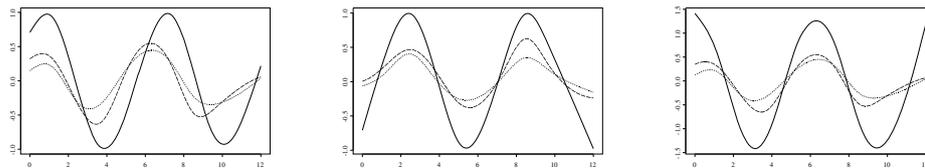*IMSE* criterium with stepwise order

CVMSE criterium with variability order

CVMSE criterium with stepwise order

RIVAR criterium with variability order

RIVAR criterium with stepwise order

Figure 7. Simulated functional parameters (solid line) and the average of their optimum estimations in the 500 simulations with the two different orders of inclusion of PCs (variability and stepwise) and the two different PCA: PCA1-type (dashed line) and PCA2-type (dotted line).

Table 3. Summary measures of correct classification rates provided by functional linear discrimination analysis in the 500 repetitions of the simulation study.

Measure	CCR
Min	32.50
1st Quartile	55.00
Median	58.75
Mean	58.76
3rd. Quartile	62.5
Max	75.00
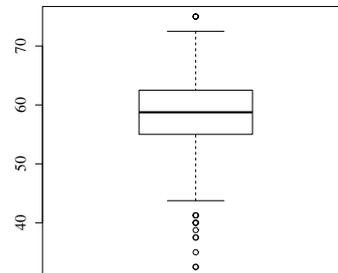
2003). In this section a base-line category logit model is used to classify near infrared (NIR) spectra of corn samples according to the spectrometer that generates them. This is a retrospective study of curves of spectrum.

The NIR spectra of 80 corn samples were measured by three different instruments at Cargill Inc. (m5, mp5 and mp6 spectrometers). The wavelength domain was $[1100, 2498]$ nm, measured at 2 nm intervals (700 observations). The data can be downloaded from the web site http://www.eigenvector.com/data/Corn/index.html. The original 80 corn samples data set was split into 60 samples for training purposes (training sample) and 20 samples for testing (test sample). These NIR spectra has been used in Tan and Brown (2003) for predicting oil content of the corn samples (multivariate calibration) and can be seen in Figure 8.

Least squares approximation on the basis of cubic B-splines with 30 equally spaced knots on the wavelength range has been used for reconstructing the sample curve of each NIR spectrum (Figure 9). A base-line category logit model is considered for estimating the categorical response that represents the measuring instrument from the NIR spectrum. The estimation of the parameters of this model will be performed by regressing the categorical response on a set of functional principal components. Given that the results provided by the two different forms of functional PCA considered in this work are very similar, only the computationally simplest form (PCA1) is presented.

From Table 4 we can observe the power of the base-line category logit model as a classification method, giving in almost all cases (except the trivial models with one PC) correct classification rates that are very close to 100%. From this table we could conclude that following the cross-validation criterium the best possible response prediction (spectrometer classification) would be provided by the model with 8 PC's included by variability order and by the model with 5 PC's included by stepwise order. These two models provide *CCR*=100% in the training sample, and *CVCCR* of 99.44% and 96.67%, respectively. The *CCR* of the test sample is 100% in variability order and 98% in stepwise order.
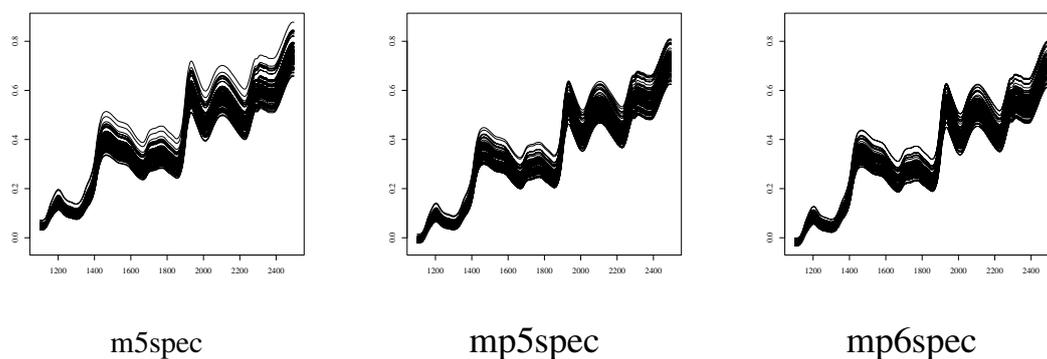
m5spec   mp5spec   mp6spec

Figure 8. Curves of corn spectrum measured with the 3 different spectrometers.
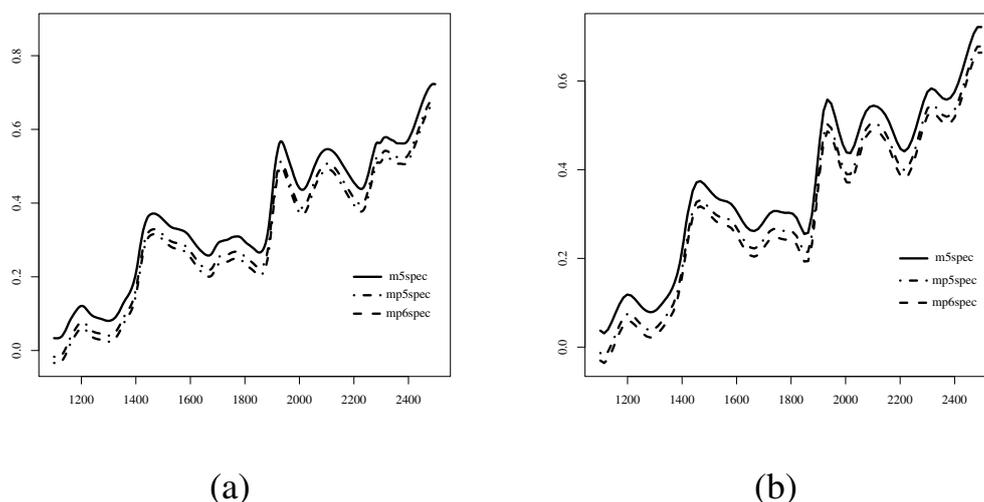

(a)   (b)

Figure 9. Observed curves of the NIR spectrum of a corn sample measured with the 3 different spectrometers (a) and their least squares approximation on the basis of cubic B-splines with 30 equally spaced knots (b).

Looking for interpretability of the functional parameters, their smoothest estimations are achieved in all cases by the model with the first two principal components (see Figure 10). This model would provide *CCR*=93.33% in the test sample and provides an interpretation of the odds of selecting one spectrometer instead other. The functional parameter beta1 is associated with the odds of mp5spec against m5spec and beta2 is associated with the odds of mp6spec against m5spec. From Figure 10 we can observe that in low wavelengths (under 1950nm) beta1 is always over beta2 with negative values and opposite in high wavelengths (over 1950nm). Moreover the form of this estimated functional parameters is very similar to the observed NIR spectra of the sample what could help the interpretation of the functional parameters. So the form of the functional parameters could suggest that the increment of NIR spectra in high wavelengths increases the odds that the

Table 4. Goodness of fit measures (*CCR* and *MSE*) of models in terms of different number of functional PCs included by variability order and stepwise order. FPCA1. Each row presents the measures of the model with as many FPCs as indicates the first column. The cumulated variance of the different number of FPCs is also included. The last two columns presents the results of the crossvalidation.

s	pc's	CumVar	CCR	MSE	CVCCR	CVMSE
			Variability order			
1	1	96.33	53.33	1.83E-1	52.78	0.1879
2	2	99.74	86.11	6.90E-2	84.44	0.0758
3	3	99.92	92.22	3.64E-2	90.00	0.0418
4	4	99.96	93.89	2.78E-2	93.33	0.0337
5	5	99.98	95.00	1.59E-2	93.89	0.0291
6	6	99.99	100.00	1.55E-6	96.67	0.0168
7	7	99.99	100.00	9.59E-12	98.89	0.0075
8	8	100.00	100.00	1.89E-11	100.00	1.2E-5
9	9	100.00	100.00	1.56E-11	99.44	0.0033
10	10	100.00	100.00	1.81E-11	99.44	0.0037
…	…	…	…	…	…	…
32	32	100.00	100.00	1.001e-011	99.44	0.0037
			Stepwise order			
1	2	3.42	63.33	0.1425	62.22	0.1468
2	1	99.74	86.11	0.0690	84.44	0.0758
3	3	99.92	92.22	0.0364	90.00	0.0418
4	6	99.93	97.22	0.0142	95.56	0.0215
5	7	99.94	100.00	4.565E-5	96.67	0.0195

spectrum is generated by mp6spec or mp5spec instead of m5spec, with this odds being higher for mp6spec than mp5spec. Just the opposite happens in low wavelengths. In other words, high NIR spectra in high wavelengths are associated with higher probability with mp6spec and mp5spec spectrometers (in this order). Meanwhile in low wavelengths the highest values of the spectrum are associated with higher probability with the reference spectrometer (m5spec).

In this example we have also obtained the correct classification rate provided by functional linear discrimination analysis, that was 87.083%, again a bit lower than the one obtained with functional multinomial logit model.

6.   Classification of the Phoneme Data

In this section we illustrate the ability of functional multinomial logit model for classification in a classical example that has been widely used in the literature. As described in Hastie, Tibshirani, and Friedman (2008), the phoneme data consist of 4509 log-periodograms of length 256 associated with five known class (phoneme) memberships (*sh*, *dcl*, *iy*, *aa* and *ao*)
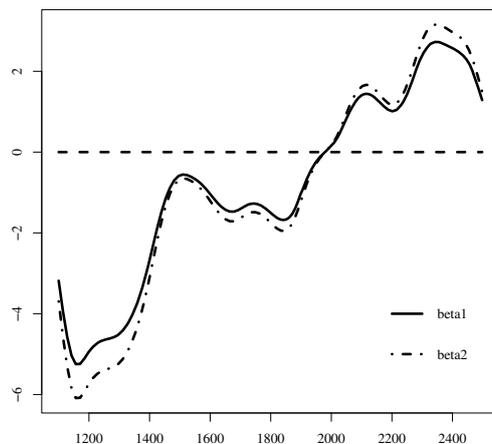
Figure 10. Estimation of the parameter functions given by the base-line category logit model in terms of the two first PCs. The function beta1 is associated with the odds of spectrometer mp5spec against m5spec and beta2 is associated with the odds of mp6spec against m5spec.

selected from continuous speech of 50 male speakers, 4509 speech frames of 32 msec duration with approximately 2 examples of each phoneme from each speaker (more details in Hastie, Tibshirani, and Friedman 2008).

The log-periodograms curves were reconstructed by least squares approximation in terms of the cubic B-splines defined the unequally spaced knots $\{0.00, 31.25, 62.50, 93.75, 125.00, 156.25, 187.50, 218.75, 250.00\}$ and five of them can be seen in Figure 11 with their associated raw data.

In this example the fitting of the functional model provides smooth functional parameters (see Figure 12) and good classification rate (93.08%) which makes it unnecessary to use functional principal components to get these estimations. This rate is very similar and a bit better than the one provided by Hastie, Tibshirani, and Friedman (2008) that was 92% by using the alternative approach given by fused LASSO (see Tibshirani, Saunders, Rosset, Zhu, and Knight 2005).

7. Conclusions

The problem of classification of curves is solved in this paper from a functional data analysis point of view that takes into account the high correlation between curves at nearby time points. A functional base-line category logit model is considered to predict the class membership to which a sample curve belongs to. This model also allows to establish the relationship between the multi-category response variable that represents the class membership and the functional predictor variable that generates the observed curves. Functional PCA and basis expansion of sample curves are proposed to estimate the functional parameters of the model. This way the

phoneme *sh*   phoneme *iy*

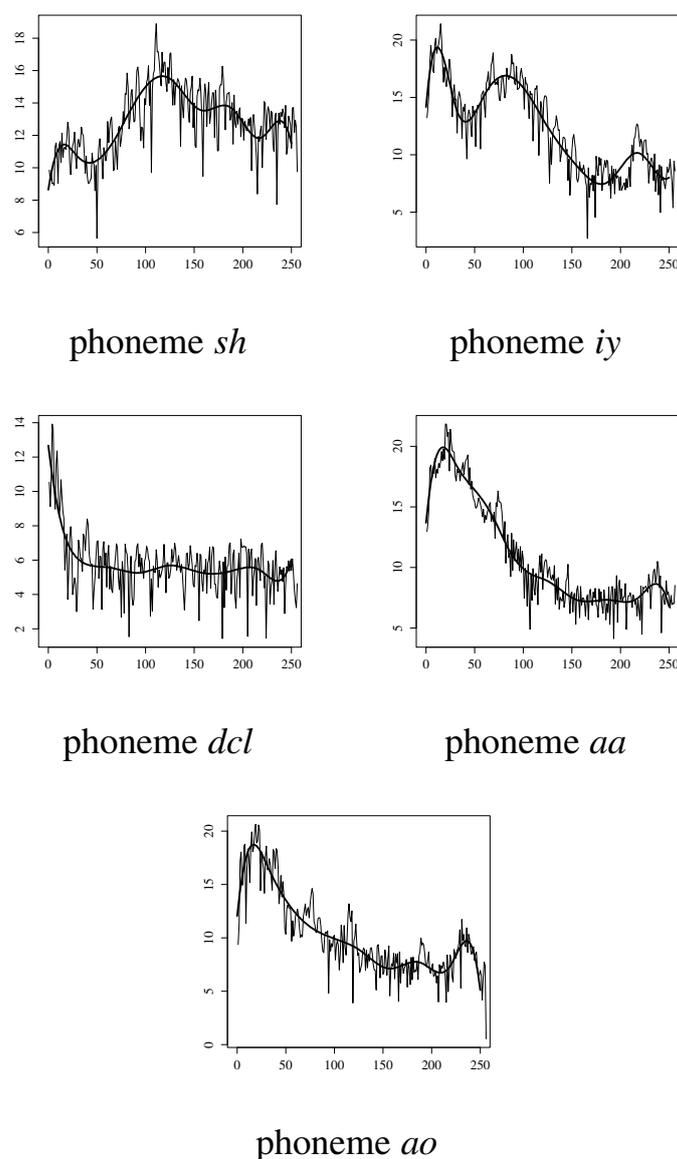phoneme *dcl*   phoneme *aa*

phoneme *ao*

Figure 11. Five selected curves and raw data of periodograms associated with the five considered phonemes in the phoneme data.

functional model is reduced to logit regression on a set of functional principal components.

The power of FPCA to estimate the parameter function of functional base-line logit models and to classify functional data is shown on a simulation study where the performance of different model selection criteria is studied. Two different FPCA approaches that agree with orthonormal basis are compared with the cubic B-splines basis. After repeating the simulation five hundred times, we can conclude that the results provided by both FPCAs are very similar. With respect to the criterium for selecting the optimum PCs, it has been shown that stepwise selection and leave-one-out cross-validation provides more parsimonious models than variability order

$$\widehat{\beta}_1(t) \qquad \widehat{\beta}_2(t)$$
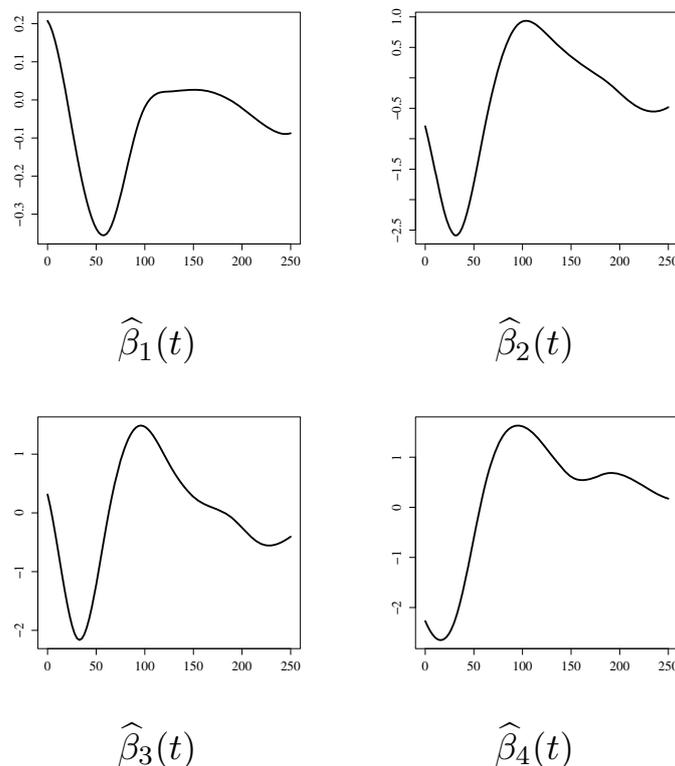
$$\widehat{\beta}_3(t) \qquad \widehat{\beta}_4(t)$$

Figure 12. Estimated functional parameters in the phoneme data example.

and the associated errors (*IMSE*, *PMSE* and *CVMSE*) are not significantly larger.

References

AGRESTI, A. (2002), *Categorical Data Analysis*, New York: Wiley.

AGUILERA, A.M., GUTIÉRREZ, R., and VALDERRAMA, M.J. (1996), "Approximation of Estimators in the PCA of a Stochastic Process Using B-splines", *Communications in Statistics - Simulation and Computation, 25(3)*, 671–690.

AGUILERA, A.M., ESCABIAS, M., and VALDERRAMA, M.J. (2006), "Using Principal Components for Estimating Logistic Regression with High-Dimensional Multicollinear Data", *Computational Statistics and Data Analysis, 50(8)*, 1905–1924.

AGUILERA, A.M., ESCABIAS, M., and VALDERRAMA, M.J. (2008), "Discussion of Different Logistic Models with Functional Data. Application to Systemic Lupus Erythematosus", *Computational Statistics and Data Analysis, 53(1)*, 151–163.

AGUILERA, A.M., ESCABIAS, M., PREDA, C., and SAPORTA, G. (2010), "Using Basis Expansions for Estimating Functional PLS Regression. Applications with Chemometric Data", *Chemometrics and Intelligent Laboratory Systems, 104(2)*, 289–305.

CARDOT, H., FAIVRE, R., and GOULARD, M. (2003), "Functional Approaches for Predicting Land Use with the Temporal Evolution of Coarse Resolution Remote Sensing Data", *Journal of Applied Statistics, 30(10)*, 1185–1199.

CARDOT, H., and SARDA, P. (2005), "Estimation in Generalized Linear Models for Functional Data Via Penalized Likelihood", *Journal of Multivariate Analysis, 92*, 24–41.

CHAMROUKHI, F., SAMÉ, A., GOVAERT, G., and AKNIN, P. (2010), "A Hidden Process Regression Model for Functional Data Description. Application to Curve Discrimination", *Neurocomputing, 73*, 1210–1221.

ESCABIAS, M., AGUILERA, A.M., and VALDERRAMA, M.J. (2004), "Principal Component Estimation of Functional Logistic Regression: Discussion of Two Different Approaches", *Journal of Nonparametric Statistics, 16(3–4)*, 365–384.

ESCABIAS, M., AGUILERA, A.M., and VALDERRAMA, M.J. (2005), "Modelling Environmental Data by Functional Principal Component Logistic Regression", *Environmetrics, 16(1)*, 95–107.

ESCABIAS, M., AGUILERA, A.M., and VALDERRAMA, M.J. (2007), "Functional PLS Logit Regression Model", *Computational Statistics and Data Analysis, 51(10)*, 4891–4902.

ESCABIAS, M., VALDERRAMA, M.J., AGUILERA, A.M., SANTOFIMIA, M. E., and AGUILERA-MORILLO, M. C. (2013), "Stepwise Selection of Functional Covariates in Forecasting Peak Levels of Olive Pollen", *Stochastic Environmental Research and Risk Assessment, 27(2)*, 367–376.

FERRATY, F., and VIEU P. (2003), "Curves Discrimination: A Nonparametric Functional Approach", *Computational Statistics and Data Analysis, 44(1-2)*, 161–173.

HASTIE, T., TIBSHIRANI, R., and FRIEDMAN, J. (2008), *The Elements of Statistical Learning. Data Mining, Inference, and Prediction,* (2nd. ed.), New York: Springer.

HERVÁS, C., SILVA, M., GUTIÉRREZ, P.A., and SERRANO, A. (2008), "Multilogistic Regression by Evolutionary Neural Network as a Classification Tool to Discriminate Highly Overlapping Signals: Qualitative Investigation of Volatile Organic Compounds in Polluted Waters by Using Headspace-Mass Spectrometric Analysis", *Chemometrics and Intelligent Laboratory Systems, 92(2)*, 179–185.

JAMES, G.M., and HASTIE, T.J. (2001), "Functional Discriminant Analysis for Irregularly Sampled Curves", *Journal of the Royal Statistical Society. Series B, 63(3)*, 533–555.

JAMES, G.M. (2002), "Generalized Linear Models with Functional Predictors", *Journal of the Royal Statistical Society, Series B, 64(3)*, 411–432.

KAYANO, M., DOZONO, K., and KONISHI, S. (2010), "Functional Cluster Analysis Via Orthonormalized Gaussian Basis Expansions and Its Application", *Journal of Classification, 27*, 211–230.

MARX, B.D., and EILERS, P.H.C. (1999), "Generalized Linear Regression on Sampled Signals and Curves. A P-spline Approach", *Technometrics, 41*, 1–13.

MASSY, W.F. (1965), "Principal Component Regression in Exploratory Statistical Research", *Journal of the American Statistical Association, 60(309)*, 234–256.

MATSUI, H., ARAKI, T., and KONISHI, S. (2011), "Multiclass Functional Discriminant Analysis and Its Application to Gesture Recognition", *Journal of Classification, 28*, 227–243.

MÜLLER, H.G., and STADTMÜLLER, U. (2005), "Generalized Functional Linear Models", *The Annals of Statistics, 33(2)*, 774–805.

OCAÑA, F.A., AGUILERA, A.M., and ESCABIAS, M. (2007), "Computational Considerations in Functional Principal Component Analysis", *Computational Statistics, 22(3)*, 449–466.

PREDA, C., SAPORTA, G., and LÉVÉDER, C. (2007), "PLS Classification of Functional Data", *Computational Statistics, 22(2)*, 223–235.

RAMSAY, J.O., and SILVERMAN, B.W. (2002), *Applied Functional Data Analysis*, New York: Springer-Verlag.

RAMSAY, J.O., and SILVERMAN, B.W. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer-Verlag.

RATCLIFFE, S.J., LEADER, L.R., and HELLER, G.Z. (2002), "Functional Data Analysis with Application to Periodically Stimulated Foetal Heart Rate Data. II: Functional Logistic Regression", *Statistics in Medicine, 21(8)*, 1115-1127.

SAEYS, W., De KETELAERE, B., and DAIRUS, P. (2008), "Potential Applications of Functional Data Analysis in Chemometrics", *Journal of Chemometrics, 22*, 335–344.

TAN, H., and BROWN, S.D. (2003), "Multivariate Calibration of Spectral Data Using Dual-Domain Regression Analysis", *Analytica Chimica Acta, 490,* 291–301.

TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J., and KNIGHT, K. (2005), "Sparsity and Smoothness Via the Fused Lasso", *Journal of the Royal Statistical Society, Series B, 67(1)*, 91–108.

VALDERRAMA, M.J., OCAÑA, F.A., AGUILERA, A.M., and OCAÑA-PEINADO, F.M. (2010), "Forecasting Pollen Concentration by a Two-Step Functional Model", *Biometrics, 66*, 135–144.