Functional PLS regression with functional response : the basis expansion approach.

Cristian Preda[1] and Jang Schiltz[2]

[1] Université des Sciences et Technologies de Lille, Laboratoire Paul Painlevé, 59650 Villeneuve d'Ascq, Cedex, France
(E-mail: `cristian.preda@polytech-lille.fr`)

[2] Université du Luxembourg, Luxembourg School of Finance, 4, rue Albert Borschette, L-1246 Luxembourg
(E-mail: `jang.schiltz@uni.lu`)

Abstract. This paper considers Partial Least Squares regression with both the predictor and the response variables of functional type. Approximating the predictor and the response in finite dimensional functional spaces, we show the equivalence of the PLS regression with functional data and the finite multivariate PLS regression applied to the expansion basis coefficients corresponding to the predictor and to the response. This equivalence is expressed in terms of PLS components and model performance. We give the explicit formula for reconstructing the coefficient regression function from the regression coefficients of the multivariate finite PLS model. A simulation study is presented.

Keywords: PLS regression, functional data, basis expansion.

1 Introduction

Partial Least Squares (PLS) is one of the most used regularisation techniques for the estimation of linear regression models. Developed in the multivariate finite dimensional setting (i.e. both predictor and response are finite dimensional) by the works of Wold et al. ([9]), it is extended by Preda and Saporta ([7]) to the case of functional data predictor $X = \{X_t\}_{t \in \mathcal{T}_X}$, $\mathcal{T}_X \subset \mathbb{R}$, and scalar response Y.

Consider $\mathcal{T}_X = [0, T]$, $T > 0$ and write the linear functional model with scalar response as

$$Y = \int_{\mathcal{T}_X} \beta(t) X_t dt + \varepsilon, \tag{1}$$

where $X = \{X_t\}_{t \in \mathcal{T}_X}$ is a real-valued stochastic process L_2-continuous with sample paths in $L_2(\mathcal{T}_X)$. Y is a real-valued random variable with finite variance and ε is the residual.

It is important to note that if the projection of Y on the linear space spanned by $\{X_t\}_{t \in \mathcal{T}_X}$ exists, in general, it can not be written as $\hat{Y} = \int_{\mathcal{T}_X} \beta(t) X_t dt$ (see [7]). Thus, the problem of estimating the coefficient function β in the linear functional model using the least square criterion is an ill-posed one (see

Cardot et al. [3]). Therefore, regularising techniques of the least square criterion are proposed in order to approximate $\hat{Y}$ by $\int_{\mathcal{T}_X} \beta(t) X_t dt$. One of the first proposed techniques is the principal components regression (PCR) where a subset of the principal components of X is used for regression. The question of the choice of principal components retained for regression is important and difficult to solve (for more details see Escabias et al. [5]). The functional PLS approach proposed in [7] consists of penalising the least squares criterion by maximising the covariance (Tucker criterion) instead of the correlation coefficient. In both techniques the main tool is the Escoufier operator W^X associated to the predictor X, defined by

$$W^X Z = \int_{\mathcal{T}_X} \mathbb{E}(X_t Z) X_t dt, \quad \forall Z \text{ r.r.v.} \tag{2}$$

Note that the spectral analysis of W^X provides the principal components of X.

In practice, the functional data X are in general not observed in all points $t \in \mathcal{T}_X$ but in a discrete way. Often, for each path $x = \{x(t), t \in \mathcal{T}_X\}$ of X (curve or trajectory) one associates a finite discrete set of time points for the observation of x, $x = \{x(t_i), i = 1, \ldots, p_x\}$. Different curves can have different time points. One way to recover the functional character of the data from the discrete observations is to consider that the sample paths of X belong to a finite dimensional subspace of $L_2(\mathcal{T}_X)$ spanned by a basis of functions $B_X = \{\phi_1, \ldots, \phi_K\}$, $K \geq 1$. Hence,

$$X(t) \approx \sum_{i=1}^{K} \alpha_i \phi_i(t), \quad \forall t \in \mathcal{T}_X, \tag{3}$$

where the α_i's are real random variables representing the coefficients associated to the approximation of X by an element of the space spanned by the basis B_X. The vector of coefficients $\alpha = \{\alpha_i\}_{i=1,\ldots K}$ can be obtained using a smoothing or interpolation methods.

It is expected that under the approximation (3), the regression models can be expressed in terms of α and the basis B_X. Let us denote by Φ the $K \times K$ matrix with entries $\Phi_{i,j} = \langle \phi_i, \phi_j \rangle_{\mathcal{T}_X}$ and by $\Phi^{\frac{1}{2}}$ the square root of Φ, i.e. $\Phi = [\Phi^{\frac{1}{2}}]^T [\Phi^{\frac{1}{2}}]$, where M^T stands for the transpose of M. Then, the Escoufier operator W^X is approximated by the Escoufier operator associated to the finite dimensional random vector $\Phi^{\frac{1}{2}}\alpha$, $W^{\Phi^{\frac{1}{2}}\alpha}$, defined by

$$W^{\Phi^{\frac{1}{2}}\alpha} Z = [\mathbb{E}(\Phi^{\frac{1}{2}}\alpha \cdot Z)]^T \Phi^{\frac{1}{2}}\alpha = \big(\mathbb{E}(\alpha_1 Z), \ldots, \mathbb{E}(\alpha_K Z)\big) \begin{bmatrix} \\ \Phi \\ \\ \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{pmatrix}. \tag{4}$$

Under the approximation (3), in Aguilera et al. ([1]) the authors show the equivalence between the PLS regressions of Y on predictors X, respectively

α. An explicit formula is also given for the coefficient regression function β of the model (1) using the PLS criterion.

The functional response. The PLS regression with both response $Y = \{Y_t, t \in \mathcal{T}_Y\}$ and predictor $X = \{X_t\}_{t \in \mathcal{T}_X}$ of functional type ($\mathcal{T}_Y \subset \mathbb{R}$) is presented in [7] (Section 3.1). The linear model is written as

$$Y(s) = \int_{\mathcal{T}_X} \beta(s,t) X(t) dt + \varepsilon_t, \quad t \in \mathcal{T}_Y, \tag{5}$$

where β is the coefficient regression function and $\{\varepsilon_t\}_{t \in T_Y}$ the residuals.

The main result is that at the each step of the PLS regression, the PLS components $t_h, h \geq 1$ are eigenvectors of the product of the two Escoufier's operators, W^X and W^Y,

$$W^X W^Y t_h = \lambda t_h. \tag{6}$$

Let consider in the following that both the response and the predictor are approximated in finite dimensional function spaces, that is

$$\begin{aligned} X(t) &\approx \sum_{i=1}^{K} \alpha_i \phi_i(t), \quad \forall t \in \mathcal{T}_X, \\ Y(t) &\approx \sum_{i=1}^{L} \gamma_i \psi_i(t), \quad \forall t \in \mathcal{T}_Y, \end{aligned} \tag{7}$$

where α_i's and ϕ_i's are as in (3) and γ_i's form a random vector γ of dimension L, $L \geq 1$, representing the coefficients associated to the approximation of Y by an element of the subspace of $L_2(\mathcal{T}_Y)$ spanned by the basis $B_Y = \{\psi_1, \dots, \psi_L\}$, $L \geq 0$.

In this paper we show the equivalence of the PLS estimate of the regression of Y on X ($PLS(Y \sim X)$) and the PLS estimate of the regression of the vector γ on the vector α ($PLS(\gamma \sim \alpha)$) with some metrics associated to the basis function B_X and B_Y. Thus, the PLS regression in the functional framework is reduced to that in the finite multivariate setting. In addition, we give an explicit formula for the coefficient regression function β of model (5), as a function of the regression coefficients obtained in $PLS(\gamma \sim \alpha)$.

The paper is organised as follows. In Section 2 we present the theoretical framework of functional PLS regression with functional response and give our main result after approximation of predictor and response into finite dimensional function spaces. In section 3, finally, we present a simulation study.

2 Functional PLS regression with functional response

The PLS approach consists of penalising the least squares criterion by maximising the covariance (Tucker criterion) instead of the correlation coefficient.

These ideas have been efficiently used in the finite dimensional case in the work of [9].

The PLS components associated to the regression of the response $Y = \{Y_t\}_{t\in\mathcal{T}_Y}$ on the functional predictor $X = \{X_t\}_{t\in\mathcal{T}_X}$, are obtained as solutions of the Tucker criterion extended to functional data as

$$\max_{\substack{w \in L_2(\mathcal{T}_X),\, \|w\|_{L_2(\mathcal{T}_X)} = 1 \\ c \in L_2(\mathcal{T}_Y),\, \|c\|_{L_2(\mathcal{T}_Y)} = 1}} Cov^2\left(\int_{\mathcal{T}_X} X_t w(t)dt, \int_{\mathcal{T}_Y} Y_t c(t)dt\right). \quad (8)$$

Let us denote by W^X, respectively W^Y, the Escoufier's operators associated to X, respectively to Y, defined by (2). Then, as it is shown in [7], the first PLS component t_1 of the regression of Y on X, is given by the eigenvector associated to the largest eigenvalue of the operator $W^X W^Y$:

$$W^X W^Y t_1 = \lambda_{\max} t_1.$$

Let $X_0 = X$ and $Y_0 = Y$. Then, the first PLS-step is completed by ordinary linear regression of X_0 and Y_0 on t_1. Let denote by X_1 and Y_1 the residuals of these linear regression models

$$\begin{aligned} X_{1,t} &= X_{0,t} - p_1(t)t_1 \quad t \in \mathcal{T}_X, \\ Y_{1,t} &= Y_{0,t} - c_1(t)t_1, \quad t \in \mathcal{T}_Y. \end{aligned} \quad (9)$$

The weight function $w_1(t)$ associated to the first PLS component t_1 is given by

$$w_1(t) = \frac{\int_{\mathcal{T}_Y} \mathbb{E}(Y_s X_t) ds}{\sqrt{\int_{\mathcal{T}_X} \left[\int_{\mathcal{T}_Y} \mathbb{E}(Y_s X_t) ds\right]^2 dt}}, \quad t \in \mathcal{T}_X,$$

so that

$$t_1 = \int_0^T w_1(t) X(t) dt.$$

At the step h, $h \geq 1$, of the PLS regression of Y on X, we define the h^{th} PLS component, t_h, by the eigenvector associated to the largest eigenvalue of the operator $W^X_{h-1} W^Y_{h-1}$

$$W^X_{h-1} W^Y_{h-1} t_h = \lambda_{\max} t_h,$$

where W^X_{h-1} and W^Y_{h-1} are the Escoufier's operators associated respectively to X_{h-1} and Y_{h-1}.

Finally, the PLS step is completed by the ordinary linear regression of X_{h-1} and Y_{h-1} on t_h. Denote by $X_h$and Y_h the functional random variables which represent the error of these regressions,

$$\begin{aligned} X_{h,t} &= X_{h-1,t} - p_h(t)t_h,\ t \in \mathcal{T}_X, \\ Y_{h,t} &= Y_{h-1,t} - c_h(t)t_h, \quad t \in \mathcal{T}_Y, \end{aligned} \quad (10)$$

The properties of the PLS components are summarised by the following proposition ([7]):

Proposition 1. For any $h \geq 1$

a) $\{t_h\}_{h\geq 1}$ *forms an orthogonal system in the linear space spanned by* $\{X_t\}_{t\in\mathcal{T}_X}$,
b) $Y_t = c_1(t)t_1 + c_2(t)t_2 + \ldots + c_h(t)t_h + Y_{h,t}, \quad t \in \mathcal{T}_Y$,
c) $X_t = p_1(t)t_1 + p_2(t)t_2 + \ldots + p_h(t)t_h + X_{h,t}, \quad t \in \mathcal{T}_X$,
d) $\mathbb{E}(Y_{h,t}t_j) = 0, \quad \forall t \in \mathcal{T}_Y, \forall j = 1, ..., h$,
e) $\mathbb{E}(X_{h,t}t_j) = 0, \quad \forall t \in \mathcal{T}_X, \forall j = 1, ..., h$.

Let us now consider the approximations of Y and X as in (7) and denote by

$$\Lambda = \Phi^{\frac{1}{2}}\alpha \quad \text{and} \quad \Pi = \Psi^{\frac{1}{2}}\gamma, \tag{11}$$

the matrix associated to the two-by-two inner-product of basis functions B_X and B_Y. Here $\Phi^{\frac{1}{2}}$ and $\Psi^{\frac{1}{2}}$ are the square roots of Φ and Ψ respectively. Notice that Φ and Ψ are symmetric real-valued matrix of size $K \times K$, respectively $L \times L$.

Our main result is :

Proposition 2.

i) The PLS regression of Y on X is equivalent to the PLS regression of Π on Λ in the sense that at each step h of the PLS algorithm, $1 \leq h \leq K$, we have the same PLS components for both regressions.
ii) If Σ is the $L \times K$-matrix of the regression coefficients of Π on Λ obtained with the PLS regression at step h, $1 \leq h \leq K$,

$$\Pi = \Sigma\Lambda + \varepsilon, \tag{12}$$

then the PLS approximation at step h of the regression coefficient function β from (5) is given by

$$\hat{\beta}(t, s) = \sum_{i}^{L}\sum_{j}^{K} S_{i,j}\psi_i(t)\phi_j(s), \quad (t, s) \in \mathcal{T}_Y \times \mathcal{T}_X, \tag{13}$$

where $S = [\Psi^{\frac{1}{2}}]^{-1}\Sigma[\Phi^{\frac{1}{2}}]^{-1}$.

The proof is based on the observation that

$$W^X \approx W^{\Phi^{\frac{1}{2}}\alpha} \quad \text{and} \quad W^Y \approx W^{\Psi^{\frac{1}{2}}\gamma}$$

and is made by induction upon h. Details are available from the authors.

This result establishes the equivalence between the PLS regression of Y on X with the PLS regression of γ on α with metrics Ψ and Φ in the sense defined by Cazes [4]. It provides a practical way to obtain the approximation of the regression coefficient function β when the functional data are approximated in finite dimensional function spaces. More precisely, it says that giving the expansion coefficients of curves representing X (α), respectively Y (γ), into the spaces spanned by B_X and B_Y, one can use them in a finite multivariate standard PLS regression (see equation (12)) in order to obtain the approximation of β. Finally, let us say that the function *plsr* of the R package `pls` can do this efficiently. It was used in our simulation and application examples.

3 Simulation study

Let us consider that X is a functional random variable defined by

$$X_t = \sum_{i=1}^{K=7} \alpha_i \phi_i(t), \quad t \in \mathcal{T}_X = [0,1], \tag{14}$$

where the $\{\alpha_i\}_{i=1,\ldots,K}$ are independent r.v.'s identically distributed uniformly on the interval $[-1;1]$ and $\phi = \{\phi_i\}_{i=1,\ldots K}$ is a cubic B-spline basis on $[0,1]$.

Let us define

$$Y(t) = \int_0^1 \beta(t,s) X_s ds + \varepsilon_t, \quad t \in \mathcal{T}_Y = [0,1], \tag{15}$$

where $\beta(t,s) = (t-s)^2$, $\forall (t,s) \in [0,1]^2$, and ε_t is the residual.

One obtains :

$$\begin{aligned} &\mathbb{E}(X_t) = 0, \qquad \mathbb{V}(X_t) = \frac{1}{3}\sum_1^K \phi_i^2(t), && \forall t \in [0,1], \\ &\mathbb{E}(Y_t) = 0, \qquad \mathbb{V}(Y_t) = \frac{1}{3}\sum_1^K \left(\int_0^1 \beta(t,s)\phi_i(s)ds\right)^2, && \forall t \in [0,1]. \end{aligned}$$

The residual $\{\varepsilon_t\}_{t\in[0,1]}$ is a zero-mean random process such that the ε_t are normally distributed with variance $\sigma_t^2 > 0$ and ε_t and ε_s are independent $\forall (s,t) \in [0,1]^2$, $s \neq t$. The residual variance, σ_t^2, is chosen such that the signal-noise ratio, $\dfrac{\mathbb{V}(Y_t)}{\mathbb{V}(\varepsilon_t)}$ is controlled. In our simulation we considered $\dfrac{\mathbb{V}(Y_t)}{\mathbb{V}(\varepsilon_t)} = 0.8, \forall t \in [0,1]$.

We simulated a sample of $n = 100$ paths of X and computed the corresponding ones for Y. The graph in Figure 1 shows such a sample.

We look for an estimation of β from the sample of curves $\{(x_i, y_i)\}_{i=1,\ldots,n}$, with $n = 100$. In order to apply the Proposition 2 we have considered for the approximation of Y a trigonometric basis of $L = 7$ functions, $\psi = \{\psi_i\}_{i=1,\ldots,L}$. The coefficients γ are then estimated by least squares regression using an equidistant grid of 100 points of $[0,1]$.

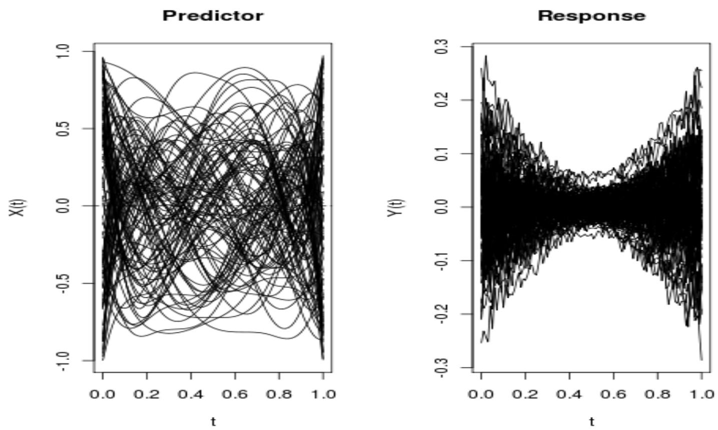

Fig. 1. A sample of n=100 paths of the response and the predictor functional variables

The results of our simulation are obtained averaging over 100 samples $\{(x_i, y_i)\}_{i=1,\ldots,n}$, $n = 100$, where SSE_Y, SSE_β and $\mathbb{V}_Y$ are defined by

$$SSE_Y = \int_0^1 \mathbb{E}(Y_t - \hat{Y}_t)^2 dt,$$
$$SSE_\beta = \int_0^1 \int_0^1 (\beta(t,s) - \hat{\beta}(t,s))^2 dtds \quad \text{and } V_Y = \int_0^1 \mathbb{V}(Y_t)dt.$$

SSE_Y is computed using the leave-one-out cross-validation, whereas SSE_β is computed from the model including all the n observations.

One obtains $\mathbb{V}(Y) \approx 0.00273$, $SSE_Y \approx 0.00038$ and $SSE_\beta \approx 0.00042$. The ratio $\dfrac{SSE_Y}{\mathbb{V}_Y} \approx 0.13919$ shows a good fit of the approximated model.

Acknowledgement

This work was supported by an AM2c grant of the Fonds National de la Recherche Luxembourg.

References

1. A. Aguilera, M. Escabias, C. Preda and G. Saporta. Using basis expansion for estimating functional PLS regression: Applications with chemometric data *Chemometrics and Intelligent Laboratory Systems*,104, 2, 289–305, 2010.
2. A. Aguilera, F. Ocaña and M. Valderrama. Forecasting time series by functional PCA. Discussion of several weighted approaches, *Computational Statistics* 14, 3, 443–467, 1999.

3.H. Cardot, F. Ferraty, P. Sarda. Functional linear model, *Statistics and Probability Letters*, 45, 11-22, 1999.
4.P. Cazes. Adaptation de la régression PLS au cas de la régression après Analyse des Correspondences Multiples. *Rev. Statist. Appl.* XLIV (4), 3560, 1997.
5.M. Escabias, A.M. Aguilera, M.J. Valderrama. Principal component estimation of functional logistic regression: discussion of two different approaches, *Journal of Nonparametric Statistics* 16(3-4), 365-384, 2004.
6.H. Lian. Nonlinear functional models for functional responses in reproducing kernel Hilbert spaces, *Canadian Journal of Statistics*, 35-4, 597–606, 2007.
7.C. Preda and G. Saporta. PLS regression on a stochastic process *Computational Statistics and Data Analysis*, 48, 149–158, 2005.
8.J.O. Ramsay B.W. Silverman. *Functional Data Analysis*, Second edition, Springer, 2005.
9.S. Wold, A. Ruhe, W.J. Dunn III. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* 5 (3), 735–743, 1984.