PLS regression on a stochastic process

Cristian Preda

CERIM - Département de Statistique
Faculté de Médecine
Université de Lille 2
1, Place de Verdun
59045 Lille Cedex
e-mail : cpreda@univ-lille2.fr

Gilbert Saporta

CNAM - Paris
Chaire de Statistique Appliquée
CEDRIC
292, Rue Saint Martin
75141 Paris Cedex 03
e-mail : saporta@cnam.fr

Abstract

We give an extension of PLS regression to the case where the set of predictor variables forms a L_2-continuous stochastic process and the response is a random vector of finite or infinite dimension. We prove the existence of PLS components as eigenvectors of some operator and also some convergence properties of the PLS approximation. The results of an application to stock-exchange data will be compared with those obtained by others methods.

Keywords. PLS regression, Escoufier's operator, principal component analysis.

1 Introduction

It doesn't seem usual to perform a linear regression when the number of predictors is infinite. However it is the case when one tries to predict a response variable Y thanks to the observation of a time dependent variable X_t, for any $t \in [0, T]$ (for example, $(X_t)_{t\in[0,T]}$ can represent temperature curves observed in n places and Y the amount of crops). Theoretically, this can be expressed by the regression of the Y variable on the $(X_t)_{t\in[0,T]}$ process.

The aim of this article is to adapt the PLS regression when all the explicative variables form a stochastic process. The problems brought about by the classical linear regression on a process – the indetermination of the regression coefficients (Ramsay and Dalzell ([10]), Ramsay and Silverman ([11]), Saporta ([12])) or the choice of the principal components of $(X_t)_{t\in[0,T]}$ as explicative variables (Deville ([4]), Saporta ([12]), Aguilera et al. ([1])) – get within this framework satisfactory solutions the main characteristics of which derive from those of the Escoufier operator associated with the process $(X_t)_{t\in[0,T]}$ (Saporta ([12])).

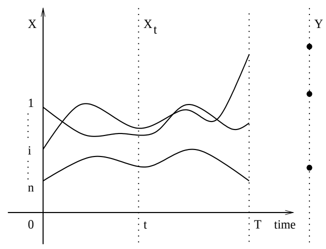

Figure 1: Regression on a stochastic process

PLS regression on a stochastic process is an extension of the finite case (a finite set of predictors) developed by Wold et al.([15]), Tenenhaus et al. ([13]) and Cazes ([2]). We prove the existence of PLS components as well as a few convergence properties towards the classical linear regression. The case $\mathbf{Y} = (X_t)_{t\in[T,T+a]}$, $a > 0$, presents an alternative to prevision methods proposed by Aguilera et al. ([1]) and Deville ([4]). The results of an application on stock exchange data are compared with those obtained by other methods.

2 Some results about PCA and regression when data are curves

Let $(X_t)_{t\in[0,T]}$ be a random process and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)$, $p \geq 1$, a random vector defined on the same probability space $(\Omega, \mathcal{A}, P)$. We assume that $(X_t)_{t\in[0,T]}$ and $\mathbf{Y}$ are of second order, $(X_t)_{t\in[0,T]}$ is L_2-continuous and for any $\omega \in \Omega$, $t \mapsto X_t(\omega)$ is an element of $L_2([0,T]$. Without loss of generality we assume also that $E(X_t)=0$, $\forall t \in [0,T]$ and $E(Y_i)=0$, $\forall i=1,\ldots,p$.

2.1 PCA of a stochastic process

Also known as Karhunen-Loève expansion, the principal component analysis (PCA) of the stochastic process $\{X_t\}_{t\in[0,T]}$ consists in representing X_t as :

$$X_t = \sum_{i\geq 1} f_i(t)\xi_i, \quad \forall t \in [0,T], \tag{1}$$

where the set $\{f_i\}_{i\geq 1}$ (the principal factors) forms an orthonormal system of deterministic functions of $L_2([0,T])$ and $\{\xi_i\}_{i\geq 1}$ (principal components) are uncorrelated zero-mean random variables.

The principal factors $\{f_i\}_{i\geq 1}$ are solution of the eigenvalue equation :

$$\int_0^T C(t,s)f_i(s)ds = \lambda_i f_i(t), \tag{2}$$

where $C(t,s)=\mathrm{cov}(X_t,X_s)$, $\forall t,s \in [0,T]$. Therefore, the principal components $\{\xi_i\}_{i\geq 1}$ are defined as $\xi_i = \int_0^T f_i(t)X_t dt$.

For $f,g \in L_2([0,T])$ let $\mathbf{C}$ be the covariance operator, $f \overset{\mathbf{C}}{\longmapsto} g$,

$$g(t) = \int_0^T C(t,s)f(s)ds.$$

Then, (2) can be written as

$$\mathbf{C}f_i = \lambda_i f_i, \quad i \geq 1.$$

Thus, the principal component analysis of $(X_t)_{t\in[0,T]}$ is the extension of the classical case corresponding to a finite number of random variables (see also ([3]). This analogy is summarized in the following table :

	Finite	**Infinite**
Variables	$X_1, \ldots, X_p,\ p \geq 1$	$X_t,\ t \in [0,T]$
Data	n vectors $\in \mathbf{R}^p$	n curves $\in L_2([0,T])$
Covariance	Matrix $\mathbf{V}$ $\mathbf{V} \in \mathcal{M}_{p\times p}(\mathbf{R})$	Operator $\mathbf{C}$ $\mathbf{C} : L_2([0,T]) \to L_2([0,T])$
Factors	Vector $u \in \mathbf{R}^p$ $\mathbf{V}u = \lambda u$	Function $f \in L_2([0,T])$ $\mathbf{C}f = \lambda f$

2.2 Linear regression on a stochastic process

Let $\mathbf{Y}$ be one-dimensional random vector ($p = 1$) and put $\mathbf{Y} = \{Y\}$. Thus, the approximation of Y given by the linear regression on the stochastic process $(X_t)_{t\in[0,T]}$ is the projection $\hat{Y}$ of Y on $L_2(X)$. By the projection theorem, $\hat{Y}$ satisfies the following system

$$E(YX_t) = E(\hat{Y}X_t), \quad \forall t \in [0,T]. \tag{3}$$

The linear model consists in writing $\hat{Y}$ as

$$\hat{Y} = \int_0^T \beta(t)X_t dt, \quad \beta \in L_2([0,T]), \tag{4}$$

and, from (3), we get that the function β must verify the Wiener-Hopf equation :

$$E(X_tY) = \int_0^T C(t,s)\beta(s)ds, \quad \forall t \in [0,T]. \tag{5}$$

The solution of this equation is caracterized by the Picard theorem ([12]) : the equation (5) has an unique solution in $L_2([0,T])$ if and only if

$$\sum_{k\geq 1} \frac{c_k^2}{\lambda_k^2} < \infty, \quad c_k = \int_0^T E(X_tY)f_k(t)dt, \tag{6}$$

where $\{\lambda_k\}_{k\geq 1}$ and $\{f_k\}_{k\geq 1}$ are the elements of the PCA of $(X_t)_{t\in[0,T]}$.

If the condition (6) is not verified, β is a distribution instead of a function ([12]). This difficulty appears also in practice when one tries to estimate the regression coefficients, $\beta(t)$, using a sample of size N. Indeed, if $\{(Y_1, X_1), (Y_2, X_2), \ldots (Y_N, X_N)\}$ is a sample of (Y, X), $X = (X_t)_{t\in[0,T]}$, the system

$$Y_i = \int_0^T X_i(t)\beta(t)dt, \quad \forall i = 1, ..., p, \tag{7}$$

has an infinite number of solutions ([11]). Several approaches are proposed ([6], [8], [11]). In the next section we present two of these.

2.2.1 Linear regression on principal components

The process $(X_t)_{t\in[0,T]}$ and the set of its principal components, $\{\xi_k\}_{k\geq 1}$, span the same linear space. Therefore, the regression of Y on $(X_t)_{t\in[0,T]}$ is equivalent to those on $\{\xi_k\}_{k\geq 1}$. Thus, we have

$$\hat{Y} = \sum_{k\geq 1} \frac{E(Y\xi_k)}{\lambda_k}\xi_k. \tag{8}$$

The fit of the regression is usually measured by :

$$R^2(Y, \hat{Y}) = \frac{1}{E(Y^2)} \sum_{k\geq 1} \frac{E^2(Y\xi_k)}{\lambda_k}.$$

In practice we need to choose an approximation of order q, $q \geq 1$:

$$\hat{Y}^q = \sum_{k=1}^{q} \frac{E(Y\xi_k)}{\lambda_k}\xi_k. \tag{9}$$

But, the choice of the principal components used to perform the regression is not easy : indeed, if the q-first principal components get in general a more stable model ([7]), the quality of this model is not necessary satisfactory according with the above criteria ([12]).

2.2.2 Projection method

Let $\{e_i\}_{i=1,\ldots,q}$ be an orthonormal system of $L_2([0,T])$. The projection method ([1], [11]) consists in approximating β with $\beta(t) \approx \sum_{i=1}^{q} b_i e_i(t)$, $b_i \in \mathbf{R}$, $\forall i = 1, \ldots q$. The relation (5) becomes

$$\mathbf{K}b = a,$$

where $\mathbf{K} = (k_{i,j})_{1 \leq i,j \leq q}$, $k_{i,j} = (\mathbf{C}e_i | e_j)$, $a = (a_i)_{i=1,\dots,p} \in \mathbf{R}^p$, $a_i = (e_i, \varphi)$, $\varphi(t) = E(YX_t), \forall t \in [0,T]$.
Notice that if $\{e_i\}_{i=1,\dots,q} = \{f_i\}_{i=1,\dots,q}$, the projection method is equivalent to the regression on the q principal components.

The choice of the approximation method employed to perform linear regression on a stochastic process seems to be a compromise between the stability of the model and its quality of fit. The PLS approach gives a solution at this compromise and is presented in the next section.

3 PLS regression on a stochastic process

Under hypothesis presented in Section 2, let $\mathbf{C}_{YX}$ and $\mathbf{C}_{XY}$ be the operators defined as :

$$\mathbf{C}_{YX} : L_2([0,T]) \to \mathbf{R}^p,\ f \xrightarrow{\mathbf{C}_{YX}} x,\ x_i = \int_0^T E(X_t Y_i) f(t) dt,\ i = 1, p,$$

$$\mathbf{C}_{XY} : \mathbf{R}^p \to L_2([0,T]),\ x \xrightarrow{\mathbf{C}_{XY}} f,\ f(t) = \sum_{i=1}^{p} E(X_t Y_i) x_i,\ t \in [0,T].$$

and denote by $\mathbf{U}_X = \mathbf{C}_{XY} \circ \mathbf{C}_{YX}$ and by $\mathbf{U}_Y = \mathbf{C}_{YX} \circ \mathbf{C}_{XY}$.

Obviously, the operators $\mathbf{C}_{YX}$ and $\mathbf{C}_{XY}$ generalize the cross-covariance matrices used in the finite case ([13]). In addition, $\mathbf{U}_X$ and $\mathbf{U}_Y$ are self-adjoint, positive and compact operators with the same spectrum. Therefore, the spectral analyses of $\mathbf{U_X}$ and $\mathbf{U_Y}$ lead to a countable set of positive eigenvalues.

The following proposition justifies the interest for these operators and gives the solution to the PLS problem :
Proposition 1 (Tucker criterion)

$$\begin{array}{c} \max \\ w, c \\ w \in L_2([0,T]), \|w\| = 1 \\ c \in \mathbf{R}^p, \|c\| = 1 \end{array} \quad Cov^2\left(\int_0^T X_t w(t) dt, \sum_{i=1}^{p} c_i Y_i\right)$$

is reached for w, respectively c, the eigenvectors associated to the largest eigenvalue of $\mathbf{U}_X$, respectively of $\mathbf{U}_Y$.

Let $w_1 \in L_2([0,T])$ be the eigenfunction of $\mathbf{U_X}$ associated to the largest eigenvalue. Then, the first PLS component (Tenenhaus et al. ([13])) of the regression of $\mathbf{Y}$ on the process $(X_t)_{t\in[0,T]}$ is the random variable defined as :

$$t_1 = \int_0^T X_t w_1(t)dt \tag{10}$$

Denote by $\mathbf{W}^X$, respectively $\mathbf{W}^Y$, the Escoufier's operators[1] (Escoufier ([5])) associated to $(X_t)_{t\in[0,T]}$, respectively to $\mathbf{Y}$, defined by :

$$\mathbf{W}^X Z = \int_0^T E(X_t Z)X_t dt, \quad \mathbf{W}^Y Z = \sum_{i=1}^{p} E(Y_i Z)Y_i, \quad \forall Z \in L_2(\Omega).$$

Our main result is the following theorem.

Theorem 2 *t_1 is the eigenvector of the $\mathbf{W}^X\mathbf{W}^Y$ associated to the largest eigenvalue.*

The PLS regression is an iterative method. Let $X_{0,t} = X_t$, $\forall t \in [0,T]$ and $Y_{0,i} = Y_i$, $\forall i = 1, ..., p$. At the step h, $h \geq 1$, of the PLS regression of $\mathbf{Y}$ on $(X_t)_{t\in[0,T]}$, we define the h^{th} PLS component, t_h, by the eigenvector associated to the largest eigenvalue of the operator $\mathbf{W}^X_{h-1}\mathbf{W}^Y_{h-1}$,

$$\mathbf{W}^X_{h-1}\mathbf{W}^Y_{h-1} t_h = \lambda_{\max} t_h, \tag{11}$$

where $\mathbf{W}^X_{h-1}$, respectively $\mathbf{W}^Y_{h-1}$, are the Escoufier's operators associated to $(X_{h-1,t})_{t\in[0,T]}$, respectively to $\mathbf{Y}_{h-1} = (Y_{h-1,i})_{i=1,...,p}$ and $\lambda_{\max}$ the largest eigenvalue of $\mathbf{W}^X_{h-1}\mathbf{W}^Y_{h-1}$. Finally, the PLS step is completed by the ordinary linear regression of $X_{h-1,t}$ and $Y_{h-1,i}$ on t_h. Let $X_{h,t}$, $t \in [0,T]$ and $Y_{h,i}$, $i = 1, ..., p$ be the random variables which represent the error of these regressions :

$$X_{h,t} = X_{h-1,t} - p_h(t)t_h, \quad t \in [0,T],$$

$$Y_{h,i} = Y_{h-1,i} - c_{h,i}t_h, \qquad i = 1, ..., p,$$

[1] The spectral analysis of this operator leads to the principal component analysis of the associated variable. See ([3]) and ([5]) for details.

As in the finite case (Tenenhaus et al. ([13])), the next statements hold :

Proposition 3 *For each $h \geq 1$:*

a) $\{t_h\}_{h\geq 1}$ *forms an orthogonal system in* $L_2(X)$*,*

b) $Y_i = c_{1,i}t_1 + c_{2,i}t_2 + \ldots + c_{h,i}t_h + Y_{h,i}, \quad i = 1, ..., p,$

c) $X_t = p_1(t)t_1 + p_2(t)t_2 + \ldots + p_h(t)t_h + X_{h,t}, \quad t \in [0,T],$

d) $E(Y_{h,i}t_j) = 0, \quad \forall i = 1, ..., p, \forall j = 1, ..., h,$

e) $E(X_{h,t}t_j) = 0, \quad \forall t \in [0,T], \forall j = 1, ..., h.$

From the Proposition 3-b), the PLS approximation of $\mathbf{Y}$ by $(X_t)_{t\in[0,T]}$ at step h, $h \geq 1$, is given by :

$$\hat{\mathbf{Y}}_h = c_1t_1 + \ldots + c_ht_h, \quad c_i \in \mathbf{R}^p, i = 1, ..., p. \tag{12}$$

Denote by $\hat{\mathbf{Y}}$ the approximation of $\mathbf{Y}$ given by the ordinary linear regression on $(X_t)_{t\in[0,T]}$. Then, the sequence $\{\hat{\mathbf{Y}}_h\}_{h\geq 1}$ is convergent in $L_2(\Omega)$ and the limit is $\hat{\mathbf{Y}}$:

Proposition 4

$$\lim_{h\to\infty} E(\|\hat{\mathbf{Y}}_h - \hat{\mathbf{Y}}\|^2) = 0. \tag{13}$$

Finally, the choice of h using the cross-validation criterion (Green and Silverman ([6])) remains applicable in this case.

Remark (Numerical solution) Because $(X_t)_{t\in[0,T]}$ is a continous time stochastic process, in practice we need a discretization of the time interval in order to get a numerical solution. In ([9]) we give such an approximation. Thus, if $\Delta = \{0 = t_0 < t_1 < \ldots < t_p = T\}$, $p \geq 1$, is a discretization of $[0,T]$, consider the process $(X_t^\Delta)_{t\in[0,T]}$ defined as :

$$X_t^\Delta = \frac{1}{t_{i+1} - t_i}\int_{t_i}^{t_{i+1}} X_t dt, \quad \forall t \in [t_i, t_{i+1}[, \quad \forall i = 0, ..., p-1.$$

Denote by m_i the random variable $\frac{1}{t_{i+1}-t_i}\int_{t_i}^{t_{i+1}} X_t dt$, $i = 0, \ldots, p-1$. Then, the approximation of the PLS regression on $(X_t)_{t\in[0,T]}$ by those on $(X_t^\Delta)_{t\in[0,T]}$ is equivalent to the PLS regression on the finite set $\{m_i\sqrt{t_{i+1} - t_i}\}$, $i = 0, \ldots, p-1$.

Continuous case Previous results are still valid for the particular case $\mathbf{Y} = (X_t)_{t\in[T,T+a]}$, $a > 0$. Indeed, because of the L_2 continuity of the process $(X_t)_{t\in[0,T+a]}$, $\mathbf{C}_{X,Y}$ and $\mathbf{C}_{Y,X}$ are compact and therefore, $\mathbf{U}_X$ and $\mathbf{U}_Y$ are compact. The results of the Proposition 1 and Theorem 2 are preserved.

The decomposition formulas (Proposition 3-b,c) become in this case :

$$X_t = \begin{cases} t_1 p_1(t) + \ldots t_h p_h(t) + X_{h,t}, & \forall t \in [0,T], \\ \\ t_1 c_1(t) + \ldots t_h c_h(t) + X_{h,t}, & \forall t \in [T, T+a], \end{cases} \tag{5}$$

For each $s \in [0,a]$, the "forecast" of X_{T+s} by $(X_t)_{t\in[0,T]}$ is given by :

$$\hat{X}_{T+s} = t_1 c_1(T+s) + \ldots + t_h c_h(T+s). \tag{6}$$

4 Application on stock exchange data

The PLS regression on a process presented in the previous sections will be used to predict the behaviour of shares on a certain lapse of time.

We have 84 shares quoted at the Paris stock exchange, for which we know the whole behavior of the growth index during one hour (between 10^{00} and 11^{00}) ; a share is likely to change every second. We also know the evolution of the growth index of a new share (noted 85) between 10^{00} and 10^{55}. The aim is to predict the way that share will behave between 10^{55} and 11^{00} using a PLS model built with the other 84 shares.

We are going to use the approximation given in Remark by taking an equidistant discretization of the interval $[0, 3600]$ (time expressed in seconds) in 60 subintervals. The previsions obtained will then match the average level of the growth index of share 85 considered on each interval $[60 \cdot (i-1), 60 \cdot i)$, $i = 56, \ldots, 60$.

Using the SIMCA-P software ([16]) we are going to build several PLS models according to the number of components chosen for regression. So we are going to refer to the model with k PLS components as PLS(k). The previsions obtained with these models will be compared to those given by the regression on the principal components (models quoted with PCR(k)) and the algorithm NIPALS (see Tenenhaus ([14]) for details). To rate the quality of those models we are going to compare the previsions obtained by each model, quoted with $\{\hat{m}\}_i(85)$, with true values $\{m\}_i(85)$, $i = 56, \ldots, 60$, observed previously.

Table 1 gives the percentages of variance associated to the three first PLS components, respectively to the three first principal components of the set $\{m_i\}_{i=1,\ldots,55}$.

PLS	%	% cum.
t_1	91.7	91.7
t_2	2.9	94.6
t_3	2.7	97.3

PCA	%	% cum.
ξ_1	91.7	91.7
ξ_2	4.0	95.7
ξ_3	2.5	98.2

Table 1: Percentages of variance

The forecasts of the some models are presented in Table 2.

	$\hat{m}_{56}(85)$	$\hat{m}_{57}(85)$	$\hat{m}_{58}(85)$	$\hat{m}_{59}(85)$	$\hat{m}_{60}(85)$	$SSE = \sum_{i=56}^{60} (\hat{m}_i - m_i)^2$
Observed	**0.700**	**0.678**	**0.659**	**0.516**	**-0.233**	-
PLS(1)	-0.327	-0.335	-0.338	-0.325	-0.302	3.789
PLS(2)	0.312	0.355	0.377	0.456	0.534	0.928
PLS(3)	0.620	0.637	0.677	0.781	0.880	1.318
PCR(1)	-0.356	-0.365	-0.368	-0.355	-0.331	4.026
PCR(2)	-0.332	-0.333	-0.335	-0.332	-0.298	3.786
PCR(3)	0.613	0.638	0.669	0.825	0.963	1.538
NIPALS	0.222	0.209	0.240	0.293	0.338	1.000

Table 2: Forecasts and errors

Considering SSE as global measure of the quality model, the best forecasts are those of the PLS regression in two steps. The models PLS(3) and PCR(3) are particularly good for short term prevision (3300, 3480), but much less for the interval of time (3480-3600). The NIPALS algorithm provides also, on an average, good predictions.

5 Conclusions

The PLS regression on a process offers an alternative to regression on principal components. It gives a solution to the problems of multicolinearity of predictors and when the number of observations is smaller than the number of explicative variables, which is often the case in this context.

References

[1] Aguilera A. M., Ocaña F., Valderama M.J. (1998) : An approximated principal component prediction model for continous-time stochastic process, Applied Stochastic Models and Data Analysis, Vol. 13, p. 61-72.

[2] Cazes P. (1997) : Adaptation de la régression PLS au cas de la régression après Analyse des Correspondances Multiples, Revue de Statistique Appliquée, XLIV (4), p. 35-60.

[3] Deville J. C. (1974) : Méthodes statistiques et numériques de l'analyse harmonique, Annales de l'INSEE, No. 15, p 3-101.

[4] Deville J. C. (1978) : Analyse et prévision des séries chronologiques multiples non stationnaires, Statistique et Analyse des Données, No. 3, p. 19-29.

[5] Escoufier Y. (1970) : Echantillonnage dans une population de variables aléatoires réelles, Publications de l'Institut de Statistique de l'Université de Paris, 19, Fasc. 4, p. 1-47.

[6] Green P.J. and Silverman B. W. (1994) : Nonparametric Regression and generalized linear models. A roughness penalty approach, Monographs on statistic and applied probability, No. 58, Chapman & Hall.

[7] Lebart L., Morineau A. and Piron, M. (1995) : Statistique exploratoire multidimensionnelle, Dunod, Paris.

[8] Palm R. and Iemma A.F. (1995) : Quelques alternatives à la régression classique dans le cas de colinéarité, Rev. Statistique Appliquée XLIII (2), p. 5-33.

[9] Preda C. (1999) : Analyse factorielle d'un processus : problèmes d'approximation et de régression, Thèse de doctorat de l'Université de Lille 1, No. 2648.

[10] Ramsay J.O. and Dalzell C.J. (1991) : Some tools for functional data analysis, Journal of Royal Statistical Society (B), 53, No. 3, p. 539-572.

[11] Ramsay J.O. and Silverman B.W. (1997) : Functional Data Analysis, Springer Series in Statistics, Springer-Verlag, New York.

[12] Saporta G. (1981) : Méthodes exploratoires d'analyse de données temporelles, Cahiers du B.U.R.O., No. 37-38, Université Pierre et Marie Curie, Paris.

[13] Tenenhaus M., Gauchi, J.P. and Menardo, C. (1995) : Régression PLS et applications, Revue de Statistique Appliquée, XLIII (1), p. 7-63.

[14] Tenenhaus M. (1998) La régression PLS. Théorie et pratique, Editions Technip, Paris.

[15] Wold S., Ruhe A., Dunn III W.J. (1984) The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, SIAM J. Sci. Stat. Comput., vol 5, no.3 pp. 735-743.

[16] Umetri AB (1996) SIMCA-P for Windows, Graphical Software for Multivariate Process Modeling, Umetri AB, Box 7960, S-90719 Umea, Sweden.