Functional PLS regression with functional response

C. PREDA[1], J. SCHILTZ[2]

[1]Ecole Polyechnique Universitaire de Lille/ INRIA Lille Nord Europe, France

[2]Luxembourg School of Finance, University of Luxembourg, Luxembourg

Introduction

`Predictor` : $X = \{X_t,\ t \in \mathcal{T}_X\} \quad X_t : \mathbf{\Omega} \to \mathbb{R}$,

`Response` : $Y = \{Y_t,\ t \in \mathcal{T}_Y\} \quad Y_t : \mathbf{\Omega} \to \mathbb{R}$.

Hypothesis :

- $E(X_t^2) < \infty$, $E(Y_t^2) < \infty$,
- X and Y are L_2–continuous,
- $\forall \omega \in \mathbf{\Omega}$:
  $\{X_t(\omega),\ t \in \mathcal{T}_X\} \in L_2(\mathcal{T}_X)$, $\{Y_t(\omega),\ t \in \mathcal{T}_Y\} \in L_2(\mathcal{T}_Y)$,
- $E(X_t) = 0$, $\forall t \in \mathcal{T}_X$, $E(Y_t) = 0$, $\forall t \in \mathcal{T}_Y$.

The linear model is written as

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

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

Least squares criterion : Wiener-Hopf equation :

$$Cov(Y(s)X(t)) = \int_{\mathcal{T}_X} Cov(X(t), X(r))\beta(s,r)dr, \;\; s \in \mathcal{T}_Y,$$

In general, no unique solution !

Solutions :

- Functional Principal Components
- Basis expansion approximation (Malfait and Ramsay (2003))
- Penalized functional linear model (Harezlak et.al (2007), Eilers and Marx (1996))

Partial Least Squares : Tucker criterion

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

$$\mathbf{C}_{XY} : L_2(\mathcal{T}_Y) \to L_2(\mathcal{T}_Y) :$$

$$\mathbf{C}_{XY}(f)(t) = g(t) = \int_{\mathcal{T}_Y} \mathbb{E}(X_t Y_s) ds, \quad f \in L_2(\mathcal{T}_Y), t \in \mathcal{T}_X$$

$$\mathbf{C}_{YX} : L_2(\mathcal{T}_X) \to L_2(\mathcal{T}_X) :$$

$$\mathbf{C}_{YX}(g)(t) = f(t) = \int_{\mathcal{T}_X} \mathbb{E}(Y_t X_s) ds, \quad g \in L_2(\mathcal{T}_X), t \in \mathcal{T}_Y$$

$$\mathbf{U}_X = \mathbf{C}_{XY} \circ \mathbf{C}_{YX} : \quad \mathbf{U}_X w = \lambda w.$$

$$\mathbf{U}_Y = \mathbf{C}_{YX} \circ \mathbf{C}_{XY} : \quad \mathbf{U}_Y c = \lambda c.$$

First PLS component :

$$t_1 = \int_{\mathcal{T}_X} X_t w(t) dt.$$

Escoufier's operators :

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

$$W^Y Z = \int_{\mathcal{T}_Y} \mathbb{E}(Y_t Z) Y_t dt, \quad \forall Z \text{ r.r.v.}$$

$$W^X W^Y t_1 = \lambda_{\max}^{(1)} t_1$$

PLS iteration :

Let $X_0 = \{X_{0,t} = X_t, \forall t \in \mathcal{T}_X\}$ and $Y_0 = \{Y_{0,t} = Y_t, \forall t \in \mathcal{T}_Y\}$.

Step h, $h \geq 1$:

$$\mathbf{W}^X_{h-1}\mathbf{W}^Y_{h-1}t_h = \lambda^{(h)}_{\max}t_h.$$

Simple linear regression on t_h :

$$X_{h,t} = X_{h-1,t} - p_h(t)t_h, \quad t \in \mathcal{T}_X,$$
$$Y_{h,t} = Y_{h-1,t} - c_h(t)t_h, \quad t \in \mathcal{T}_Y.$$

Proposition.

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) $\mathbb{E}(X_{h,t}t_j) = 0, \quad \forall t \in \mathcal{T}_X, \forall j = 1, ..., h.$

$$\hat{Y}_t = c_1(t)t_1 + c_2(t)t_2 + \ldots + c_h(t)t_h = \int_{\mathcal{T}_X} \beta_{PLS(h)}(t,s)X(s)ds.$$

PLS and basis expansion.

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

Metrics :

$\Phi = \{\phi_{i,j}\}_{1\leq i,j\leq K}$, $\phi_{i,j} = \langle \phi_i, \phi_j \rangle_{L_2(\mathcal{T}_X)}$

$\Psi = \{\psi_{i,j}\}_{1\leq i,j\leq L}$, $\psi_{i,j} = \langle \psi_i, \psi_j \rangle_{L_2(\mathcal{T}_Y)}$

Coefficients :

$\Lambda = \Phi^{1/2}\alpha, \quad \Pi = \Psi^{1/2}\gamma$

Equivalence with finite dimensional PLS regression

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, $1 \leq h \leq K$,

$$\Pi = \Sigma\Lambda + \varepsilon,$$

then the PLS approximation at step h of the regression coefficient function β is given by

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

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

Simulation study

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

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 $[0,1]$ with equidistant knots.
Let define

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

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

One obtains :

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

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 $\dfrac{\mathbb{V}(Y_t)}{\mathbb{V}(\varepsilon_t)} = 0.9, \forall t \in [0,1]$.

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

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

From the cross-validation scores obtained for each response in the vector Π, $h = 5$ seems a good choice for the number of PLS components. The model obtained with $h = 5$ PLS components gives the following matrix S,

$$S = \begin{pmatrix} -0.0551 & 1.4171 & 1.7687 & 1.5224 & 0.2926 & -0.9046 & -0.6530 \\ 0.0559 & 0.4597 & 1.2733 & 0.8860 & 0.0741 & -0.5878 & -0.5195 \\ 0.0262 & -0.5313 & 0.4191 & -0.0219 & -0.2213 & -0.0678 & -0.2143 \\ 0.8245 & 1.1117 & -0.9000 & 0.3107 & -0.2631 & -0.1077 & 0.2155 \\ 1.1491 & 2.3503 & -1.4923 & 0.6788 & -0.2762 & -0.2187 & 0.4060 \end{pmatrix}.$$

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.

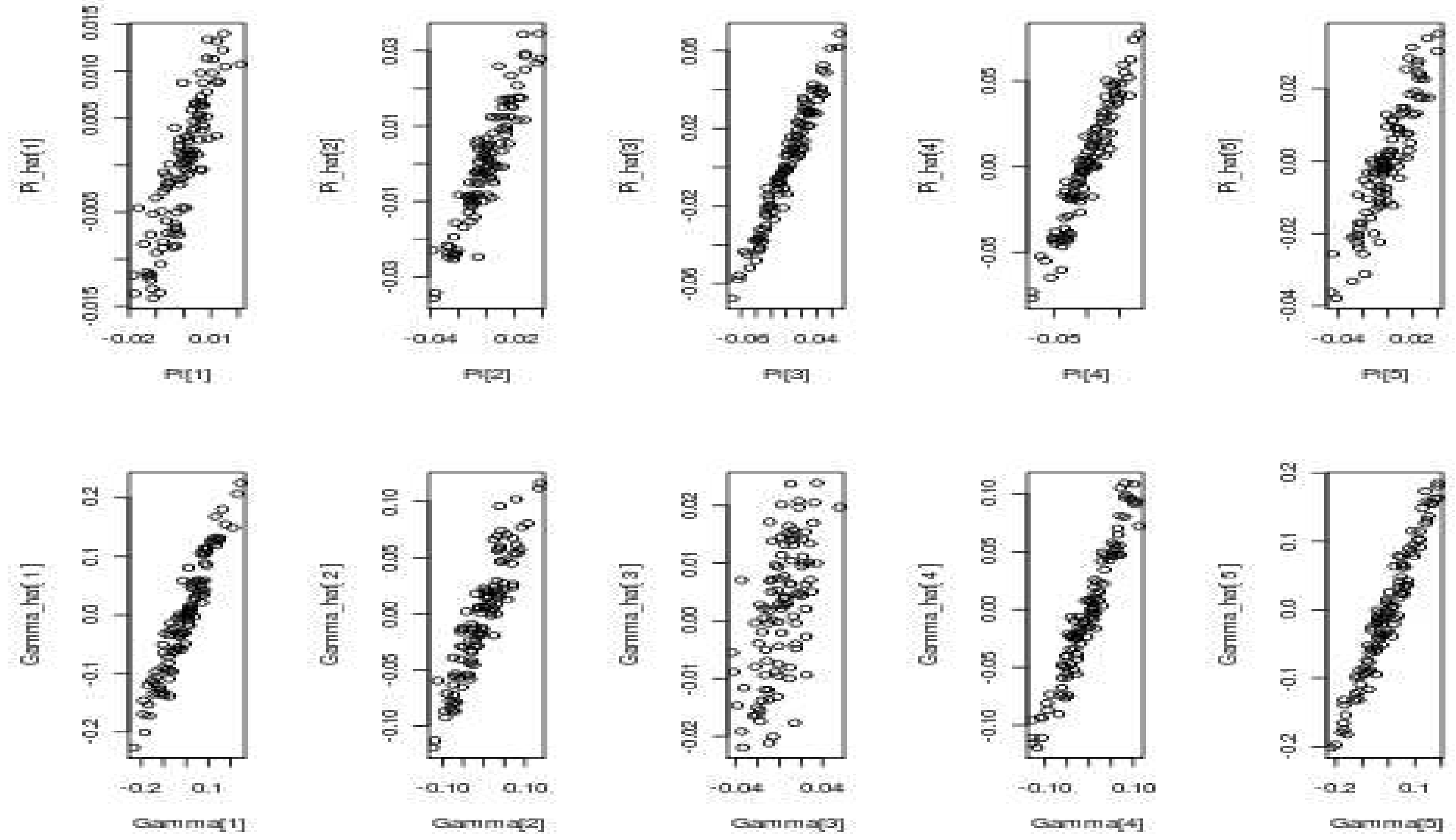

Figure 1: Predicted and observed values for the expansion coefficients γ and Π