Theoretical Comparison of the Functional Principal Component Analysis and Functional Partial Least Squares

Marine Carrasco   Idriss Tsafack

Université de Montréal University of California Irvine

Montreal

October 2021

NBER-NSF Time series conference

Introduction

- We consider a simple functional model with a functional predictor $X(t)$ and a scalar response Y.

$$Y_i = \alpha + \int_S \beta(s) X_i(s) ds + \varepsilon_i,$$

where t can be time or spatial location or any other index and ε_i is the error term. We observe (Y_i, X_i) $i = 1, ..., n$.

- The object of interest is the estimation of the slope function $\beta(s)$.

- The main problem: we deal with an **ill-posed problem** because of the dimensionality.

- Some regularization (smoothing) is needed.

Partial Least Squares (PLS):

Wold (1975), Helland (1988), Preda and Saporta (2005), Reiss and Ogden (2007), Aguilera et al. (2010), Delaigle and Hall (2012), Febrero-Bande et al. (2015), Blanchard and Krämer (2016), Blazère, Gamboa, Loubes (2014). PLS in econometrics : Groen and Kapetanios (2008), Kelly and Pruitt (2015), Carrasco and Rossi (2016).

- PLS looks at combinations of the independent variables which are highly correlated with the dependent variable. PLS is popular in the multivariate linear regression when the number of predictor variables is larger than the number of observations. It is also useful when there is a multicolinearity issue.

- The theoretical comparison of PCA and PLS is not yet developed for the functional regression.

- So far, the comparison was based on Monte-Carlo Simulations. Preda and Saporta (2005), Delaigle and Hall (2012), Febrero-Bande et al. (2015).

The model

$$Y_i = \int_S \beta(s)X_i(s)ds + \varepsilon_i. \tag{1}$$

- $(X_i, Y_i); i \in \{1, ..., n\}$ a sequence of i.i.d random variables with the same distribution as (X, Y).
- We assume that X and Y are zero-mean.
- $(\varepsilon_i; i \in \{1, ..., n\})$ is a sequence of i.i.d random error with $E(\varepsilon_i|X) = 0$ and $E(\varepsilon_i^2|X) = \sigma^2$.
- $L^2(S)$ the space of all square-integrable functions mapping from on a compact interval S to $\mathbb{R}$. The inner product $< ., . >$ and norm $||.||$ are respectively defined as $< f, g > = \int_S f(s)g(s)dt$ and $||f|| = \left(\int_S f^2\right)^{1/2}$.

Premultiplying by $X(t)$ and taking the expectation on the both sides of the model equation lead to the following result.

$$\mathbb{E}[X(t)Y] = \int_S \mathbb{E}[X(t)X(s)]\beta(s)ds + \mathbb{E}[X(t)\varepsilon] \qquad (2)$$

Since $E(\varepsilon_t|X) = 0$, we obtain that

$$C_{xy} = K\beta$$

where K is the covariance operator, $K : L^2(S) \to L^2(S)$

$$Kf = \mathbb{E}[(X \otimes X)(f)] \qquad (3)$$

C_{xy} denotes the cross-covariance function, $C_{xy} : S \to \mathbb{R}$

$$C_{xy}(t) = \mathbb{E}[X(t)Y] \qquad (4)$$

We cannot solve directly

$$\beta = K^{-1}C_{xy}$$

because K is not invertible in the space $L^2(S)$. See Carrasco, Florens, and Renault (2007).

Functional Partial Least Squares

Standard PLS

$$\min_{\beta} \| \underline{Y} - \langle \underline{X}, \beta \rangle \|_n^2$$

$$\text{s.t. } \beta \in \mathcal{K}_m \left(\widehat{C}_{xy}, \widehat{K} \right) = span \left(\widehat{C}_{xy}, \widehat{K}\widehat{C}_{xy}, ..., \widehat{K}^{m-1}\widehat{C}_{xy} \right).$$

Modified PLS

$$\min_{\beta} \left\| \widehat{C}_{xy} - \widehat{K}\beta \right\|^2$$

$$\text{s.t. } \beta \in \mathcal{K}_m \left(\widehat{C}_{xy}, \widehat{K} \right) = span \left(\widehat{C}_{xy}, \widehat{K}\widehat{C}_{xy}, ..., \widehat{K}^{m-1}\widehat{C}_{xy} \right).$$

The modified PLS estimator is

$$\hat{\beta}_m^{PLS} = \sum_{j=0}^{m-1} \widehat{\gamma}_j \widehat{K}^j \widehat{C}_{xy}$$

where

$$\hat{\gamma} = H^{-1} a$$

and H is a $m \times m$ Hankel matrix and a is a $m \times 1$ vector with

$$(H)_{ij} = \underline{Y}' W_n^{i+j+1} \underline{Y} \text{ and } a_i = \underline{Y}' W_n^{i+1} \underline{Y}$$

and W_n is a $n \times n$ matrix with element $\langle X_i, X_j \rangle / n$.

Remark 1. $\widehat{\gamma}_j$ depend nonlinearly on $Y \Rightarrow \hat{\beta}_m^{PLS}$ is **nonlinear** in Y.

Remark 2. For PCA, $\hat{\beta}_m^{PCA}$ is **linear** in Y.

Assumptions

A1. (X_i, Y_i) are i.i.d with the same distribution law as (X,Y).

A2. $\int_S \beta^2(t)dt < +\infty$, $\int_S X^2(t)dt < +\infty$, $E[||X||^4] < \infty$,
$E[\varepsilon_i|X] = 0$, $E[\varepsilon_i^2|X] = \sigma^2$, $E[\varepsilon_i^4|X] < +\infty$

A3. We assume that β satisfies $\left\| K^{-\frac{\mu}{2}}\beta \right\|^2 = \sum_{j=1}^{\infty} \frac{<\beta,v_j>^2}{\lambda_j^\mu} < \infty$
with $\mu \geq 0$.

A4. The eigenvalues of the covariance operator K and the ones of
the empirical covariance $\hat{K}$ are distinct.
$\lambda_1 > \lambda_2 > ... > 0$ and $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_m$.

- We study the convergence rate of

$$\left\| \hat{\beta}_m^{PCA} - \beta \right\| \text{ and } \left\| \hat{\beta}_m^{PLS} - \beta \right\|.$$

- This can be decomposed as an estimation error and a regularization bias:

$$\left\| \hat{\beta}_m - \beta \right\| \leq \left\| \hat{\beta}_m - \beta_m \right\| + \left\| \beta_m - \beta \right\|.$$

- We show that, for the same number of factors m, the regularization bias for PLS is smaller than that of PCA.

$$\left\| \beta_m^{PLS} - \beta \right\| \leq \left\| \beta_m^{PCA} - \beta \right\|.$$

Rate for functional PLS

Theorem 2.

Under Assumptions A1 - A4 and given the stopping rule

$$\left\|\widehat{C}_{xy} - \widehat{K}\widehat{\beta}_m^{PLS}\right\| \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}} < \left\|\widehat{C}_{xy} - \widehat{K}\widehat{\beta}_{m-1}^{PLS}\right\|$$

where $\tau > 1$ and $A_0 = \sigma^2 \int_S E[X_i^2(s)]ds$, the rate of convergence is

$$||\widehat{\beta}_m^{PLS} - \beta||^2 = O_p\left(n^{-\mu/(\mu+2)}\right).$$

Remarks.

- This rate is **optimal** and coincides with that obtained for PCA.
- Moreover, the optimal number of components is **smaller** for PLS than for PCA.

Simulations

$Y_i = \int X_i(s) \beta(s) \, ds + \varepsilon_i$ where ε_i i.i.d. $\mathcal{N}(0,1)$ and
$X(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} u_j v_j(t)$ where $u_1, u_2, ...$ i.i.d $\mathcal{N}(0,1)$. $n = 1000$.

Model 1 (Cardot et al. 1999)
$v_j(t) = \sqrt{2} \sin((j - 0.5)\pi t)$ and $\lambda_j = \frac{1}{(j-0.5)^2 \pi^2}$, for $j = 1, 2, ...$
and $\beta(t) = 2\sin(0.5\pi t) + 4\sin(1.5\pi t) + 5\sin(2.5\pi t)$. The slope is a linear function of exactly 3 eigenfunctions.

Model 5 (Hall and Horowitz, 2007)
$v_1(t) = 1$, $\lambda_1 = 1$, $v_j(t) = \sqrt{2}\cos((j-1)\pi t)$ and $\lambda_j = j^{-2}$,
$j = 2, 3, ...$ and $\beta(t) = \sum_{j=1}^{\infty} \beta_j v_j(t)$ with $\beta_1 = 0.3$ and
$\beta_j = 4(-1)^{j+1} j^{-2}$, for $j = 2, 3, ...$

Model 6
Same as Model 5 with $\beta_1 = \beta_2 = 0$, $\beta_3 = 3$, and
$\beta_j = 4(-1)^{j+1} j^{-2}$, for $j = 5, 6, ...$

Comparison of MSE

	Model 1		Model 5		Model 6	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	20.51	17.61	0.61	0.45	34.02	24.30
2	12.33	4.15	0.37	0.23	28.54	15.50
3	0.17	0.40	0.27	0.25	27.81	4.67
4	0.29	8.49	0.23	0.90	11.07	3.45
5	0.44	31.53	0.23	3.34	6.98	2.99
6	0.60	60.33	0.24	8.11	5.73	3.89
7	1.28	111.52	0.28	14.21	4.58	8.05
8	2.38	171.16	0.32	21.60	4.14	15.17
9	3.25	269.53	0.43	28.17	3.64	23.72
10	3.72	381.58	0.55	38.03	3.54	33.12

Comparison of Bias

	Model 1		Model 5		Model 6	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	20.50	17.60	0.61	0.45	34.01	24.23
2	12.28	4.06	0.36	0.22	28.50	15.43
3	0.02	0.04	0.25	0.11	27.71	4.42
4	0.01	1.39	0.19	0.08	10.77	3.04
5	0.04	4.13	0.15	0.09	6.71	2.18
6	0.06	8.38	0.12	0.08	5.45	1.53
7	0.04	9.18	0.10	0.15	4.24	1.27
8	0.20	19.57	0.09	0.27	3.77	1.12
9	0.10	25.29	0.08	0.20	3.17	0.98
10	0.34	41.60	0.07	0.32	2.92	1.01

Comparison of variance

	Model 1		Model 5		Model 6	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	0.01	0.01	0.01	0.01	0.01	0.07
2	0.05	0.09	0.01	0.02	0.04	0.07
3	0.15	0.36	0.02	0.14	0.09	0.25
4	0.28	7.10	0.04	0.82	0.30	0.41
5	0.40	27.41	0.08	3.25	0.27	0.81
6	0.54	51.95	0.12	8.03	0.28	2.35
7	1.24	102.34	0.18	14.06	0.34	6.78
8	2.19	151.59	0.23	21.34	0.37	14.05
9	3.14	244.24	0.35	27.97	0.48	22.75
10	3.38	339.98	0.48	37.71	0.62	32.12

Conclusion

- FPLS is an attractive alternative to FPCA.
- Both methods are rate optimal.
- The necessary number of components is in general smaller for PLS than for PCA.