Developments in Applied Statistics
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 19, Ljubljana: FDV, 2003

Selection Model in Functional Linear Regression Models for Scalar Response

Aldo Goia[1]

Abstract

The so-called Functional Linear Regression model consists in explaining a scalar response by a regressor which is a random function observed on a compact subset of $\mathbb{R}$: in this context, the "parameter" of linear model is a function of the weights.

In order to estimate this functional coefficient some estimators such as Functional Principal Component Regression Estimator, Smooth Principal Component Regression Estimator, Penalized B-Splines Estimator, have been introduced in literature. We focus our attention on the Functional Principal Component Regression Estimator and in particular on the connected dimensionality problem.

Our aim is to apply and compare some different selection methods, which have been proposed in the classical regression field. These methods are illustrated and compared by the means of simulations.

1 Introduction

In many statistical frameworks we have to face problems involving data that are curves, or more exactly, collections of discrete observations effected on curves: these data are classified as "functionals". For example we can think of longitudinal data analyzed in meteorology, in medicine and biology, in economy and/or finance, or the spectrometric data observed in chemometrics, or the digitalized signals analyzed in the signals analysis.

Many techniques for the analysis of data with functional nature which allows the variability exploration in samples of curves and let the research of common structures among the curves or groups of them have been introduced; for an exhaustive presentation of these methods, see the monograph of Ramsay and Silverman, 1997.

A very interesting application consists in the possibility of constructing regression models through which we describe the relation between a real variable and an explanatory variable having functional nature. Such a model is termed as functional regression model. Several examples can be found in literature: Hastie and Mallows

[1] Department of Economic Science and Quantitative Methods - University "A. Avogadro", Via Perrone, 18 - 28100 Novara (Italy); goia@eco.unipmn.it

(1993) propose some application to quantitative chemistry, Marx and Eilers (1996) illustrate a phonemes classification by means of the log-spectra of a sequence of spoken syllables, Ramsay and Silverman (1997) describe the estimation of the total annual precipitation in some Canadian weather station from the temperature curves measured during the year, Cardot *et al.* (1999b) propose a forecasting model in order to explain winter wheat yield as a linear function of the duration of the crop and climatic variations, Ferraty and Vieu (2002) illustrate the prevision of fat content in some meet samples from the respective spectrometric curves.

Formally in the functional regression model the link between a real random variable (r.v.) Y and a functional random variable $X = \{X(t),\ t \in \mathcal{T}\}$, where $\mathcal{T}$ is the interval of observation, is described by the relation

$$Y = \mu + \Psi(X) + \varepsilon \tag{1.1}$$

where μ is a real constant, Ψ is a real operator and ε is a zero mean random variable with finite variance, which we assume non-correlated with X.

In this paper we consider the special case where X is a r.v. mapping in $H = L^2(\mathcal{T})$, the separable Hilbert space of square integrable functions defined on $\mathcal{T} \subset \mathbb{R}$, $\mathcal{T}$ compact. In this framework Ψ is a real-valued linear continuous operator. By the Riesz Representation Theorem, there exists a unique function $\psi \in H$ such that

$$\Psi(x) = \int_{\mathcal{T}} \psi(t)x(t)\mathrm{d}t, \qquad x \in H, \tag{1.2}$$

Hence model (1.1) may be rewritten as

$$Y = \mu + \int_{\mathcal{T}} \psi(t)X(t)\mathrm{d}t + \varepsilon. \tag{1.3}$$

Model (1.3) is a generalization of the linear regression model to the case of an infinite number of regressors. Many authors dwell upon the estimation of the functional coefficient ψ and/or the operator Ψ: in some cases criteria are based on functional bases expansion of the ψ function. We mention for example the Penalized Splines Estimators studied in Marx and Eilers (1999) and in Cardot *et al.* (1999b and 2002), and the Fourier Basis Expansion Estimators treated in Ramsay and Silverman (1997).

In this work we use the estimation method introduced by Cardot *et al.* (1999a) and based on Principal Component Regression (PCR). This technique consists in a mean square estimation of regression models whose regressors are the k principal components of X, obtained from the k eigenfunctions associated to the k largest eigenvalues of the covariance operator of X. An important problem to solve is the determination of the parameter k that represents the dimension of the subspace on which we project the observations.

In order to introduce the notations we will use in the following and to clarify the model and the estimator, we devote Section 2 to a quick presentation of the functional linear model and to the discussion of the above mentioned Functional PCR Estimator.

In Section 3 we dwell upon to some selection criteria that have been proposed in the classical regression tools. These methods are applied to the problem of the optimal dimension choice and they are compared through simulation in Section 4.

2 Functional Linear Regression Model and Functional PCR Estimator

Let $H = L^2_{\mathcal{T}}$ be the Hilbert space of square integrable functions defined on a compact set $\mathcal{T} \subset \mathbb{R}$, provided of the inner product $< \cdot, \cdot >$ defined for all $f, g \in H$ as $< f, g >= \int_{\mathcal{T}} f(t)g(t)\mathrm{d}t$.

We consider a r.v. (X, Y) defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values into the measurable space $(\mathbb{R} \times H,\ \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_H)$. We assume moreover that $Y \in L^2_{\mathbb{R}}$ and $X \in L^2_H$.

Let $m(\cdot)$ be the conditional mean of Y given X, *i.e.*

$$m(x) = \mathbb{E}\left(Y | \{X(t) = x(t),\ t \in \mathcal{T}\}\right), \qquad x \in H.$$

When it is linear we can write

$$m(x) = \mu + < \psi, x >= \mu + \Psi(x), \qquad x \in H.$$

In general, since we cannot assume a linear specification for $m(.)$ we consider an approximation of it, which may be obtained as a solution of the least square problem

$$(\mu, \psi) = \arg \min_{(\alpha,\xi)\in\mathbb{R}\times H} \mathbb{E}\left[(Y - \alpha - < \xi, X >)^2\right].$$

Existence and unicity of the solution are discussed in Cardot *et al.* (2003).

The covariance operator of X and the cross-covariance operator between X and Y are defined respectively as:

$$\begin{aligned} \Gamma(x) &= \mathbb{E}\left[< (X - \mathbb{E}(X)), x > (X - \mathbb{E}(X))\right], \qquad x \in H \\ \Delta(x) &= \mathbb{E}\left[< (X - \mathbb{E}(X)), x > (Y - \mathbb{E}(Y))\right], \qquad x \in H \end{aligned}$$

These two operators are related by $\Delta = \Psi\Gamma$. Since the operator Γ is nuclear (Dauxois *et al.*, 1982), its inverse is not bounded when H has infinite dimension: to solve this drawback the idea proposed at first by Bosq in the ARH models (Bosq, 1991 and 2000) and used in Cardot *et al.* (1999a), consists of projecting X's observations on a finite $k-$dimensional subspace H_k of H.

The basic tool is the Karhunen-Loeve decomposition of a second order random function:

$$X = \sum_{j=1}^{\infty} < X, v_j > v_j = \sum_{j=1}^{\infty} \widetilde{X}_j v_j$$

where v_j are the orthonormal eigenfunctions of Γ, sorted according to the decreasing sequence of associated eigenvalues λ_j. We note that this decomposition, truncated at order $k < \infty$, gives the best linear representation of X in the sense that it maximizes the explained variance: in practice it is a Principal Component Analysis, which we call "Functional". In this context the real r.v. $\widetilde{X}_j$ plays the role of the $j-$th principal component of X.

As proven by Cardot *et al.* (2003), the operator Ψ has the following approximation

$$\Psi_{\mathrm{PCR}}(.) = \sum_{j=1}^{k} \frac{\Delta v_j}{\lambda_j} < v_j, \cdot >$$

and equivalently we have

$$\psi_{\text{PCR}} = \sum_{j=1}^{k} \frac{\Delta v_j}{\lambda_j} v_j.$$

In order to estimate Ψ (or equivalently ψ) we consider an i.i.d. sample $(X_i, Y_i)_{i=1,\dots,n}$ drawn from the r.v. (X, Y).

Define

$$\begin{aligned} \Gamma_n(x) &= \frac{1}{n}\sum_{i=1}^{n} < (X_i - \overline{X}), x > (X_i - \overline{X}), \qquad x \in H \\ \Delta_n(x) &= \frac{1}{n}\sum_{i=1}^{n} < (X_i - \overline{X}), x > (Y_i - \overline{Y}), \qquad x \in H \end{aligned}$$

the empirical versions of Γ and Δ, where $\overline{X} = \{n^{-1}\sum_{i=1}^{n} X_i(t),\ t \in \mathcal{T}\}$ and $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$.

Let $\{\widehat{v}_j, \widehat{\lambda}_j\}_{j=1,2,\dots}$ be the eigenelements of Γ_n, with $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_n = 0$, and let $\mathbb{P}_k$ be the orthogonal projection on $\widehat{H}_k$, the space spanned by $\{\widehat{v}_j\}_{j=1,\dots,k}$. The PCR estimators of Ψ and ψ are given respectively by

$$\widehat{\Psi}_{\text{PCR}}(.) = \Delta_n \mathbb{P}_k (\mathbb{P}_k \Gamma_n \mathbb{P}_k)^{-1} = \sum_{j=1}^{k} \frac{\Delta_n \widehat{v}_j}{\widehat{\lambda}_j} < \widehat{v}_j, . >$$

and

$$\widehat{\psi}_{\text{PCR}} = \sum_{j=1}^{k} \frac{\Delta_n \widehat{v}_j}{\widehat{\lambda}_j} \widehat{v}_j.$$

Thus this method consists in an ordinary least square regression of the response r.v Y on the k variables $\widetilde{X}_j$, the first k principal components of X. Under suitable conditions on k and on the spectrum of Γ and Γ_n, Cardot *et al.* (1999a) have proven the almost sure (and in probability) convergence for the PCR estimator.

As Cardot *et al.* (2003) point out, the technique gives a rough estimation of the function ψ, also when sample size n is large. In order to solve this problem the authors introduce a second step in the estimation procedure, consisting in a smoothing of $\widehat{\psi}$ by means of a B-spline approximation.

3 Model selection criteria

As we have seen in the previous section, in order to construct PCR estimator, the functional linear regression model is approximated in an optimal way by using a truncated Functional Principal Component Analysis. So we may write:

$$\begin{aligned} Y &= \mu + \Psi(X) + \varepsilon \cong \mu + \Psi\left(\sum_{j=1}^{k} \widetilde{X}_j v_j\right) + \varepsilon \\ &= \mu + \sum_{j=1}^{k} \widetilde{X}_j \Psi(v_j) + \varepsilon = \mu + \sum_{j=1}^{k} \widetilde{X}_q \beta_j + \varepsilon, \end{aligned}$$

where the eigenfunctions v_j may be estimated by $\widehat{v}_j$. This is a regression model whose explanatory variables are the first k principal components of X, ordered according to the sequence of eigenvalues λ_j, that are the variances of principal componenents (*i.e.* $\mathrm{Var}\,(\widetilde{X}_j) = \lambda_j$).

In order to construct the estimator, we need to determine of the optimal dimension k: this is a problem of variables choice in a linear regression model, a very important subject as the wide literature on the topic shows (for some analysis and discussions about selection methods, see, for example, the monographs of Miller, 1990, McQuarrie and Tsai, 1998, and Burnham and Anderson, 2002). In this work we focus our attention on some choice criteria widely used in the empirical analysis.

We consider the problem of model selection into a family of potential candidates, indexed by $h = 1, 2, \ldots$. Let $\theta_h = (\mu, \beta_1, \beta_2, \ldots, \beta_{k_h}, \sigma^2)$ be the vector of parameters of the $h-$th regression model, and $\Theta_h \subset \mathbb{R}^{k_h+2}$ the parameter space for θ_h. Denote by $d_h = k_h + 2$ the dimension of Θ_h and by $p_h = k_h + 1$ the number of coefficients in the candidate regression model.

We recall that in the PCR Estimation framework it is often suggested to order the principal component according to the magnitude of correlation with the explained variable Y, since the principal components with largest variability (i.e. associated to the largest eigenvalues) are not necessarily the most explanatory.

In our case, we can not adopt this approach since it could give some non convergent estimators. In fact, one of the hypotheses ensuring the a.s. (and in probability) convergence of PCR estimators is that sequences of eigenvalues $\{\lambda_j\}_{j=1,2,\ldots}$ and $\{\widehat{\lambda}_j\}_{j=1,2,\ldots}$ decrease (Cardot *et al.*, 1999a), and so principal components must be ranged consequently.

So, for $h = 1$, the parameter is $\theta_1 = (\mu, \sigma^2)$; for $h = 2$ we introduce the first principal component $\widetilde{X}_1$ and we have to estimate $\theta_2 = (\mu, \beta_1, \sigma^2)$. When $h = 3$ we use the first two principal components $\widetilde{X}_1$ and $\widetilde{X}_2$, and so $\theta_3 = (\mu, \beta_1, \beta_2, \sigma^2)$. The generic θ_h is constructed according to this rule.

In order to select the "good" regression model, some authors suggest to choose model h which minimizes the quantity

$$-2\log(L(h)) + pen(h) \tag{3.1}$$

where $\log(L(h))$ is the log-likelihood computed for $h-$model and $pen(h)$ is a penalty term connected to the dimension of models. Under gaussian conditions for the noise variable ($\varepsilon \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 < \infty$) criteria (3.1) may be expressed by the operative formula

$$n\log(\widehat{\sigma}_h^2) + \delta(h) \tag{3.2}$$

where $\widehat{\sigma}_h^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i^{(h)})^2$ is the ML estimator of σ^2, and $\widehat{Y}_i^{(h)}$ is the estimation of Y_i obtained by using the $h-$model.

Of this nature are the criteria introduced by Schwartz (1978) called Bayesian Information Criterion (BIC), in which $\delta(h) = d_h \log n$ and the Akaike's Information Criterion - AIC, (Akaike, 1973). In this work, we do not use the original AIC with $\delta(h) = 2d_h$, instead we make use of the "corrected" version AICc proposed by Hurvich and Tsai (1989) in which $\delta(h) = n(n+p_h)/(n-p_h-2)$ with $n > p_h+2$. This

choice is motivated by the results of Cavanaugh (1997) which proposes an unified vision of AIC and AICc.

Similar to the previous methods are the criteria proposed by Hannan and Quinn (1979) with $\delta(h) = 2d_h \log(\log n)$, and the so-called Rice's T (Rice, 1984), where $\delta(h) = n\log(1 - 2p_h/n)$ with $p_h < n/2$. The latter has been originally introduced in the context of choice of bandwidth in the kernel regression framework and here it has been adapted to the parametric case.

An other criterion that we consider is the one proposed by Shibata (1981) for which the optimal p_h minimizes the statistic

$$S(h) = \widehat{\sigma}_h^2(n + 2p_h).$$

This criterion has been proposed for some gaussian model that can be specified by an infinite number of parameters and so it may be used validly in our context.

The last approach we have considered is the classical Cross Validation consisting of setting aside of data, usually one observation at a time, and making predictions of these data by using the remaining ones. Let $\widehat{Y}_i^{(h,-i)}$ be the prediction of Y_i based on the other $n-1$ observations. We choose a p_h minimizing

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i^{(h,-i)})^2.$$

Since we treat linear models, it may be proved that it is equivalent to choose the number of coefficients p_h which minimizes the following *PRESS* statistic (Allen, 1974):

$$PRESS(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i^{(h)})^2/(1 - \Pi_i)^2$$

where Π_i is the $i-$th term of the principal diagonal of the projection matrix onto the p_h-dimensional subspace of $\mathbb{R}^n$.

4 Numerical examples

This section is devoted to numerical simulations of the performances of the criteria presented in Section 3: we make comparisons by using some different functional coefficients with various sample sizes.

In practice we compare the relative empirical *MISE* calculated for each criterion, defined as follows:

$$RMISE^{(c)} = \frac{1}{M}\sum_{m=1}^{M}\frac{\int_{\mathcal{T}}(\psi(t) - \widehat{\psi}_m^{(c)}(t))^2 \mathrm{d}t}{\int_{\mathcal{T}}(\psi(t))^2 \mathrm{d}t}$$

where M is the number of simulations in each study case and $\widehat{\psi}_m^{(c)}(t)$ is the estimation of the functional coefficient obtained at the $m-$th simulation applying the $c-$th selection method.

In our study we base simulations on the following operative conditions:

1. X is a standard brownian motion defined on the set $\mathcal{T} = [0, 1]$ and whose trajectories are discretized in 100 equispaced points,

2. sample sizes are $n = 50$, 100, 200, 500 and 1000,

3. parameters of functional linear regression models are

   (a) $\mu = 2$ for all models,

   (b) functional coefficients defined on $[0, 1]$ as follows

   i. $\psi_1(t) = \sin(.5\pi t) + .5\sin(1.5\pi t) + .25\sin(2.5\pi t)$,
   ii. $\psi_2(t) = \sin(4\pi t)$,
   iii. $\psi_3(t) = \sin(6\pi t)$,
   iv. $\psi_4(t) = (\sin(2\pi t^2))^3$,
   v. $\psi_5(t) = 1 - 48\,t + 218\,t^2 - 315\,t^3 + 145\,t^4$,
   vi. $\psi_6(t) = 0.3\exp(-64(t - .25)^2) + 0.7\exp(-256(t - .75)^2)$.

4. For each model we introduce noise variables which have gaussian, uniform and centered gamma distribution. Their variability is controlled by the following signal-to-noise ratio:

$$snr = \frac{\sigma^2_{\Psi(X)}}{\sigma^2_\varepsilon + \sigma^2_{\Psi(X)}},$$

   where $\sigma^2_{\Psi(X)} = \mathrm{Var}\,(\Psi(X))$ and $\sigma^2_\varepsilon = \mathrm{Var}\,(\varepsilon)$. In simulations we use $srn = 65\%$, 85% and 95%, which allow respectively to some error variables with large variance ($\sigma^2_\varepsilon/\sigma^2_{\Psi(X)} = 0.54$), moderate variance ($\sigma^2_\varepsilon/\sigma^2_{\Psi(X)} = 0.18$) and small variance ($\sigma^2_\varepsilon/\sigma^2_{\Psi(X)} = 0.05$).

5. The number of simulations for each setting of experimental factors is $M = 500$.

Since the eigenelements of the covariance operator for a standard brownian motion on $[0, 1]$ are (see Ash and Gardner, 1975):

$$\lambda_j = \frac{1}{(j - 0.5)^2\pi^2}, \quad v_j(t) = \sqrt{2}\sin((j - 0.5)\pi t), \ t \in [0, 1] \ \ j = 1, 2, \ldots,$$

we may observe that the functional coefficient ψ_1 is built matching the first three eigenfunctions v_j in an opportune way.

ψ_2 and ψ_3 are periodic functions which can not be decomposed into a finite sum of eigenfunctions, the second one having a large amount of fine structure. ψ_4 is more general. ψ_5 is a function which combines a less fine structure with a trend. ψ_6 has different degrees of curvature for different values of t. The shape of the functional coefficients are shown in Figures 1, 2, and 3.

Tables 1, 2, 3, 4, 5, and 6 give $RMISE$ values computed for the different functional coefficients ψ_1, ψ_2, ψ_3, ψ_4, ψ_5 and ψ_6, under various error distribution hypotheses and using the selection criteria BIC (Schwartz), AICc (Hurvich and Tsai), HQ (Hannan and Quinn), S (Shibata), T (Rice) and PRESS (Allen).

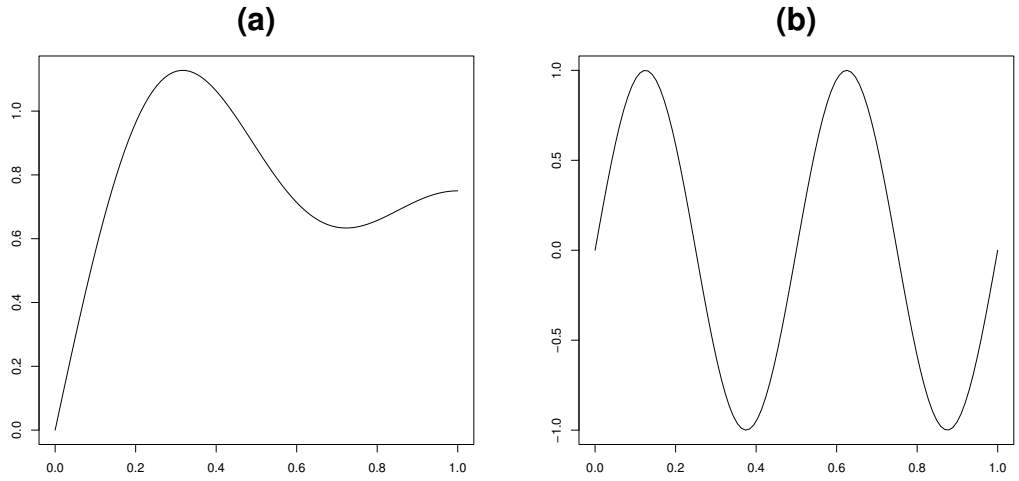

Figure 1: Functional coefficients for simulations: (a) ψ_1, and (b) ψ_2.

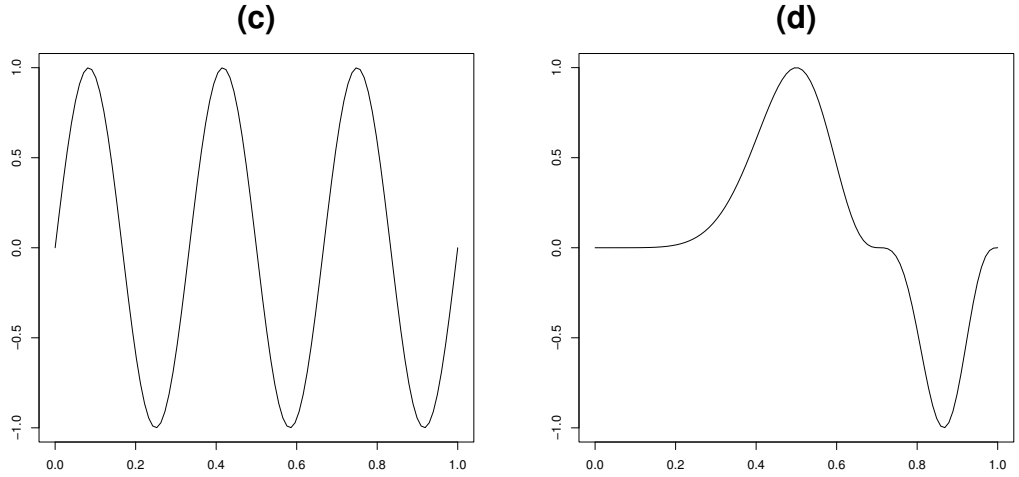

Figure 2: Functional coefficients for simulations: (c) ψ_3 and (d) ψ_4.

As general observation, we can note that the properties of the selectors differ for the various cases, but certain patterns emerge anyway, and in particular they are linked to the sample sizes, the signal-to-noise ratio (*snr* in the follows) and to the distribution of the noise variables.

It is clear that when the sample size increases, the performances of selectors improve, since estimations are more accurate. In very small samples, with $n = 50$, all approaches are unacceptable also when *snr* is large ($snr = 0.95$). However, in some cases, we obtain very large $RMISE$ values also when $n \geq 500$ (for example, see results in Table 6). Besides in general the criteria seem not to be equivalent also for very large values of n ($n = 1000$). Analogously, selectors have different behaviors

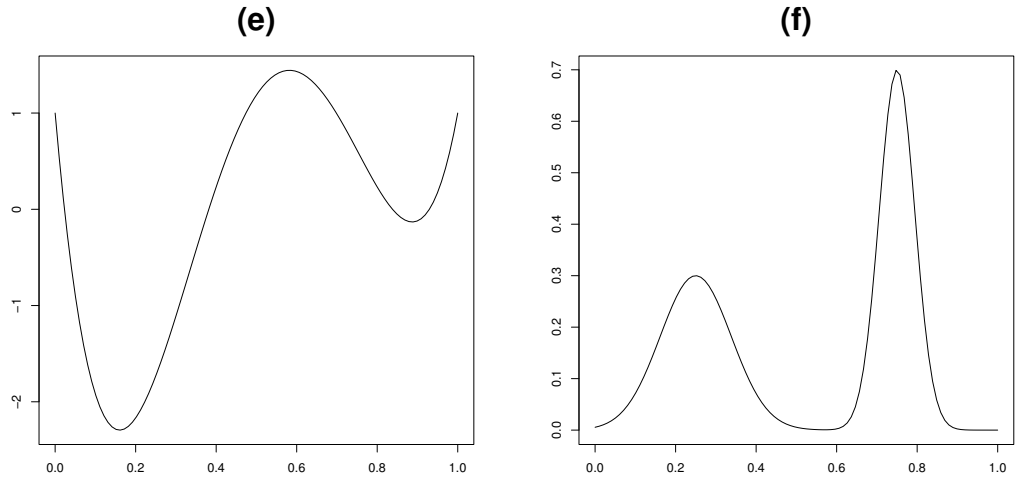

Figure 3: Functional coefficients for simulations: (e) ψ_5 and (f) ψ_6.

when *snr* change: for each sample size, $RMISE$ is reduced substantially when *snr* increases.

Another general remark is that for any case, each selector tends to reproduce the same behavior, for any error distribution. Thus, comments that may be done for the gaussian cases can be repeated changing the distributions of the error variable, both when the distribution law is symmetric (uniform case) and both when it is asymmetric (centered gamma case). This may justify the use of methods defined in gaussian context in more general cases.

In the following we propose a short discussion for each estimated model, in order to supply some practical indication for the use of selection methods.

When we estimate the functional coefficient ψ_1, BIC gives good results, constantly for any distribution error and *snr* value. For example, with $snr = 0.85$ and $n \geq 100$ $RMISE$ is smaller than 10%. The Shibata criterion is similar and sometimes better than BIC when the sample size is large (for $n = 500$ and $snr = 0.85$, we have $RMISE = 2.6\%$ for the gaussian case and $RMISE = 2.7\%$ for the other cases). The other criteria are not optimal, and in particular the one of Hannan and Quinn behaves very badly. Anyway all these criteria seem to have the same behavior for $n \geq 500$, although they give worse performances compared to BIC and S: for example, for $n \geq 500$ and $snr = 0.85$, they supply $RMISE$ that are $4 - 6$ times bigger than the best ones.

Cases of estimate of periodic functions ψ_2 and ψ_3 are rather similar with respect to the behavior of the selectors: for $n \geq 200$ the Shibata's criterion is good with $RMISE \leq 6\%$ for $snr = 0.85$, and $RMISE \leq 3.5\%$ when $snr = 0.95$, consistently for any noise distribution. Hannan and Quinn criterion perform usually in a poor way even if, for very large sample size and $snr \geq 0.85$, it seems equivalent to the other criteria.

The case of coefficient ψ_4 confirms that the Shibata's criterion is a good method of selection for large sample under different error conditions and *snr* values: in fact,

Table 1: $\psi_1(t) = \sin((1-.5)\pi t) + .5\sin((2-.5)\pi t) + .25\sin((3-.5)\pi t)$.

	n	BIC	AICc	HQ	S	T	PRESS
Gaussian Error							
snr = 0.65	50	0.3051	0.7514	9.1796	0.7544	0.8137	1.4909
	100	0.1791	0.9235	2.8973	0.3775	1.0898	1.7138
	200	0.1030	0.6501	1.2088	0.1988	0.6820	0.8321
	500	0.0569	0.2901	0.3690	0.0703	0.2909	0.3097
	1000	0.0421	0.1525	0.1726	0.0359	0.1525	0.1556
snr = 0.85	50	0.1795	0.4083	3.5345	0.5232	0.4416	0.7730
	100	0.0749	0.3531	1.1757	0.1482	0.3814	0.5520
	200	0.0537	0.2341	0.4274	0.0811	0.2502	0.3106
	500	0.0282	0.0938	0.1154	0.0256	0.0940	0.0952
	1000	0.0123	0.0674	0.0731	0.0121	0.0675	0.0678
snr = 0.95	50	0.0640	0.1201	1.3176	0.1789	0.1432	0.3127
	100	0.0406	0.1073	0.3701	0.0733	0.1180	0.1844
	200	0.0207	0.0815	0.1386	0.0239	0.0863	0.1037
	500	0.0055	0.0328	0.0395	0.0074	0.0339	0.0370
	1000	0.0025	0.0163	0.0187	0.0036	0.0163	0.0173
Centered Gamma Error							
snr = 0.65	50	0.3006	0.8081	10.6547	1.3322	0.8793	2.4219
	100	0.1840	0.9484	3.1345	0.3849	1.0845	1.2942
	200	0.1105	0.5858	1.0809	0.2064	0.6042	0.6842
	500	0.0540	0.3359	0.4534	0.0624	0.3358	0.3882
	1000	0.0419	0.1649	0.1814	0.0427	0.1648	0.1757
snr = 0.85	50	0.2612	0.3865	3.3807	0.4651	0.4296	0.8452
	100	0.0860	0.4034	1.0718	0.1857	0.4318	0.5794
	200	0.0508	0.2632	0.4835	0.0937	0.2852	0.3110
	500	0.0292	0.0854	0.1103	0.0273	0.0854	0.0910
	1000	0.0121	0.0594	0.0656	0.0123	0.0594	0.0622
snr = 0.95	50	0.0689	0.1211	1.1191	0.1709	0.1335	0.2593
	100	0.0408	0.1325	0.4054	0.0762	0.1332	0.1661
	200	0.0225	0.0793	0.1234	0.0230	0.0819	0.0895
	500	0.0058	0.0289	0.0327	0.0076	0.0290	0.0302
	1000	0.0026	0.0203	0.0229	0.0045	0.0203	0.0208
Uniform Error							
snr = 0.65	50	0.3377	0.9470	10.5057	0.9449	1.0260	1.5423
	100	0.2226	0.6962	2.9884	0.3444	0.7351	1.2375
	200	0.1145	0.5396	0.9475	0.1971	0.5889	0.6844
	500	0.0570	0.2505	0.3224	0.0649	0.2575	0.2776
	1000	0.0424	0.1556	0.1639	0.0382	0.1559	0.1539
snr = 0.85	50	0.2275	0.4183	3.8087	0.7557	0.4927	1.1520
	100	0.0907	0.3585	1.1977	0.1609	0.4235	0.6175
	200	0.0652	0.2285	0.3593	0.0766	0.2321	0.2903
	500	0.0306	0.0995	0.1311	0.0266	0.1039	0.1097
	1000	0.0130	0.0684	0.0768	0.0112	0.0685	0.0660
snr = 0.95	50	0.0875	0.1443	1.1126	0.1896	0.1574	0.3003
	100	0.0383	0.1147	0.3480	0.0631	0.1213	0.1516
	200	0.0194	0.0751	0.1255	0.0237	0.0754	0.0843
	500	0.0050	0.0376	0.0482	0.0076	0.0377	0.0378
	1000	0.0024	0.0176	0.0204	0.0038	0.0190	0.0186

Table 2: $\psi_2(t) = \sin(4\pi t)$.

	n	BIC	AICc	HQ	S	T	PRESS
Gaussian Error							
snr=0.65	50	0.3957	0.4316	1.0495	0.5045	0.4331	0.5687
	100	0.1775	0.2500	0.4171	0.2157	0.2574	0.3030
	200	0.0995	0.1338	0.1718	0.1058	0.1349	0.1412
	500	0.0521	0.0685	0.0733	0.0500	0.0685	0.0697
	1000	0.0336	0.0435	0.0451	0.0326	0.0436	0.0439
snr=0.85	50	0.2247	0.2252	0.4487	0.2655	0.2264	0.2828
	100	0.0982	0.1217	0.1794	0.1098	0.1222	0.1371
	200	0.0552	0.0683	0.0774	0.0562	0.0683	0.0714
	500	0.0301	0.0329	0.0340	0.0280	0.0329	0.0334
	1000	0.0206	0.0195	0.0199	0.0184	0.0195	0.0196
snr=0.95	50	0.1537	0.1539	0.2192	0.1685	0.1544	0.1726
	100	0.0633	0.0683	0.0867	0.0649	0.0689	0.0734
	200	0.0338	0.0364	0.0395	0.0327	0.0366	0.0372
	500	0.0177	0.0169	0.0172	0.0165	0.0169	0.0171
	1000	0.0115	0.0103	0.0104	0.0103	0.0103	0.0103
Centered Gamma Error							
snr=0.65	50	0.3598	0.3644	1.0799	0.4365	0.3696	0.4901
	100	0.1731	0.2419	0.4215	0.2132	0.2486	0.2854
	200	0.0990	0.1452	0.1971	0.1054	0.1477	0.1571
	500	0.0532	0.0687	0.0753	0.0513	0.0695	0.0704
	1000	0.0333	0.0397	0.0414	0.0315	0.0401	0.0401
snr=0.85	50	0.2266	0.2332	0.4776	0.2848	0.2343	0.2965
	100	0.0995	0.1250	0.1901	0.1087	0.1276	0.1395
	200	0.0564	0.0746	0.0898	0.0571	0.0761	0.0788
	500	0.0304	0.0333	0.0346	0.0286	0.0333	0.0334
	1000	0.0210	0.0199	0.0203	0.0189	0.0199	0.0200
snr=0.95	50	0.1504	0.1522	0.2183	0.1644	0.1508	0.1665
	100	0.0612	0.0652	0.0841	0.0625	0.0656	0.0702
	200	0.0333	0.0374	0.0409	0.0336	0.0374	0.0384
	500	0.0175	0.0174	0.0178	0.0162	0.0174	0.0175
	1000	0.0115	0.0103	0.0104	0.0102	0.0103	0.0102
Uniform Error							
snr=0.65	50	0.3639	0.3723	1.0344	0.4607	0.3784	0.5307
	100	0.1686	0.2311	0.3984	0.1857	0.2381	0.2785
	200	0.0991	0.1325	0.1749	0.0965	0.1367	0.1457
	500	0.0518	0.0645	0.0716	0.0479	0.0644	0.0681
	1000	0.0342	0.0395	0.0404	0.0317	0.0396	0.0398
snr=0.85	50	0.2198	0.2285	0.4528	0.2666	0.2272	0.2814
	100	0.0983	0.1158	0.1669	0.1089	0.1192	0.1267
	200	0.0553	0.0680	0.0833	0.0564	0.0687	0.0734
	500	0.0292	0.0345	0.0364	0.0276	0.0347	0.0352
	1000	0.0202	0.0193	0.0196	0.0179	0.0194	0.0195
snr=0.95	50	0.1562	0.1606	0.2374	0.1728	0.1599	0.1822
	100	0.0645	0.0697	0.0869	0.0668	0.0702	0.0757
	200	0.0336	0.0367	0.0405	0.0326	0.0370	0.0378
	500	0.0176	0.0171	0.0176	0.0162	0.0171	0.0172
	1000	0.0114	0.0101	0.0102	0.0102	0.0101	0.0101

Table 3: $\psi_3(t) = \sin(6\pi t)$.

	n	BIC	AICc	HQ	S	T	PRESS
Gaussian Error							
snr=0.65	50	0.5418	0.4906	0.6664	0.5245	0.4851	0.5209
	100	0.2254	0.2449	0.3089	0.2306	0.2475	0.2592
	200	0.1187	0.1341	0.1469	0.1170	0.1359	0.1377
	500	0.0618	0.0628	0.0647	0.0588	0.0628	0.0631
	1000	0.0393	0.0380	0.0385	0.0361	0.0381	0.0382
snr=0.85	50	0.2865	0.2823	0.3258	0.2869	0.2803	0.2902
	100	0.1340	0.1374	0.1518	0.1377	0.1370	0.1408
	200	0.0706	0.0716	0.0738	0.0692	0.0717	0.0722
	500	0.0362	0.0317	0.0320	0.0329	0.0317	0.0318
	1000	0.0230	0.0192	0.0191	0.0199	0.0192	0.0191
snr=0.95	50	0.1983	0.1960	0.1920	0.1933	0.1961	0.1948
	100	0.0857	0.0832	0.0847	0.0831	0.0830	0.0835
	200	0.0455	0.0414	0.0412	0.0422	0.0414	0.0414
	500	0.0218	0.0182	0.0180	0.0196	0.0182	0.0181
	1000	0.0135	0.0107	0.0106	0.0117	0.0107	0.0107
Centered Gamma Error							
snr=0.65	50	0.5378	0.4795	0.6981	0.5355	0.4754	0.5243
	100	0.2253	0.2470	0.3085	0.2351	0.2479	0.2571
	200	0.1212	0.1366	0.1517	0.1200	0.1369	0.1394
	500	0.0603	0.0644	0.0661	0.0565	0.0644	0.0644
	1000	0.0396	0.0369	0.0370	0.0363	0.0369	0.0371
snr=0.85	50	0.2972	0.2861	0.3339	0.3008	0.2845	0.2986
	100	0.1304	0.1373	0.1518	0.1329	0.1377	0.1400
	200	0.0731	0.0755	0.0779	0.0723	0.0756	0.0763
	500	0.0355	0.0323	0.0324	0.0327	0.0323	0.0324
	1000	0.0231	0.0192	0.0191	0.0202	0.0192	0.0190
snr=0.95	50	0.1951	0.1899	0.1863	0.1868	0.1894	0.1869
	100	0.0828	0.0816	0.0824	0.0819	0.0817	0.0817
	200	0.0455	0.0416	0.0413	0.0426	0.0416	0.0414
	500	0.0209	0.0181	0.0180	0.0192	0.0181	0.0181
	1000	0.0136	0.0109	0.0108	0.0118	0.0109	0.0109
Uniform Error							
snr=0.65	50	0.5301	0.4673	0.6754	0.5156	0.4661	0.5073
	100	0.2219	0.2430	0.3024	0.2282	0.2445	0.2557
	200	0.1195	0.1320	0.1469	0.1177	0.1320	0.1357
	500	0.0596	0.0611	0.0629	0.0563	0.0612	0.0614
	1000	0.0389	0.0367	0.0369	0.0346	0.0368	0.0368
snr=0.85	50	0.2958	0.2890	0.3281	0.2913	0.2853	0.2906
	100	0.1287	0.1299	0.1467	0.1284	0.1300	0.1350
	200	0.0720	0.0716	0.0747	0.0690	0.0716	0.0725
	500	0.0355	0.0341	0.0343	0.0332	0.0342	0.0341
	1000	0.0233	0.0191	0.0190	0.0203	0.0191	0.0191
snr=0.95	50	0.1990	0.1953	0.1922	0.1940	0.1951	0.1956
	100	0.0851	0.0833	0.0839	0.0834	0.0832	0.0829
	200	0.0455	0.0424	0.0420	0.0435	0.0424	0.0422
	500	0.0218	0.0188	0.0185	0.0198	0.0187	0.0186
	1000	0.0134	0.0107	0.0106	0.0116	0.0107	0.0107

Table 4: $\psi_4(t) = \big(\sin(2\pi t^2)\big)^3$.

	n	BIC	AICc	HQ	S	T	PRESS
Gaussian Error							
snr=0.65	50	0.3840	0.4621	1.5207	0.5364	0.4690	0.6364
	100	0.2710	0.3556	0.6475	0.3111	0.3572	0.4110
	200	0.2079	0.2529	0.3326	0.2116	0.2566	0.2722
	500	0.1430	0.1564	0.1602	0.1380	0.1573	0.1568
	1000	0.1188	0.0892	0.0896	0.1018	0.0892	0.0895
snr=0.85	50	0.2667	0.2868	0.6534	0.3209	0.2869	0.3643
	100	0.1805	0.2142	0.3077	0.1899	0.2157	0.2308
	200	0.1376	0.1396	0.1535	0.1333	0.1400	0.1419
	500	0.0966	0.0599	0.0614	0.0635	0.0599	0.0603
	1000	0.0480	0.0321	0.0327	0.0300	0.0322	0.0323
snr=0.95	50	0.1897	0.1943	0.3146	0.2090	0.1947	0.2227
	100	0.1154	0.1104	0.1334	0.1093	0.1110	0.1165
	200	0.0677	0.0558	0.0614	0.0541	0.0557	0.0569
	500	0.0237	0.0248	0.0251	0.0221	0.0248	0.0249
	1000	0.0142	0.0145	0.0146	0.0137	0.0145	0.0145
Centered Gamma Error							
snr=0.65	50	0.3766	0.4026	1.5415	0.5222	0.4289	0.6315
	100	0.2716	0.3786	0.6567	0.3138	0.3840	0.4268
	200	0.2029	0.2602	0.3607	0.2169	0.2644	0.2735
	500	0.1429	0.1551	0.1585	0.1391	0.1564	0.1564
	1000	0.1172	0.0870	0.0893	0.0966	0.0872	0.0881
snr=0.85	50	0.2563	0.2878	0.7081	0.3336	0.2910	0.3614
	100	0.1802	0.2236	0.3113	0.1979	0.2249	0.2408
	200	0.1393	0.1492	0.1654	0.1349	0.1482	0.1524
	500	0.0924	0.0621	0.0635	0.0660	0.0620	0.0621
	1000	0.0497	0.0335	0.0340	0.0301	0.0335	0.0338
snr=0.95	50	0.1872	0.1909	0.2986	0.2153	0.1906	0.2198
	100	0.1178	0.1124	0.1342	0.1134	0.1131	0.1166
	200	0.0692	0.0592	0.0653	0.0585	0.0595	0.0620
	500	0.0252	0.0251	0.0258	0.0229	0.0252	0.0254
	1000	0.0142	0.0148	0.0151	0.0134	0.0149	0.0148
Uniform Error							
snr=0.65	50	0.3663	0.4251	1.5492	0.5513	0.4458	0.6496
	100	0.2628	0.3704	0.6321	0.2954	0.3733	0.4414
	200	0.2011	0.2599	0.3238	0.2032	0.2627	0.2751
	500	0.1412	0.1562	0.1654	0.1401	0.1566	0.1593
	1000	0.1178	0.0885	0.0904	0.0984	0.0885	0.0885
snr=0.85	50	0.2662	0.2963	0.6546	0.3366	0.3020	0.3617
	100	0.1812	0.1994	0.2821	0.1933	0.2027	0.2240
	200	0.1360	0.1413	0.1596	0.1316	0.1416	0.1461
	500	0.0959	0.0641	0.0663	0.0668	0.0642	0.0647
	1000	0.0451	0.0325	0.0333	0.0294	0.0325	0.0327
snr=0.95	50	0.1968	0.2015	0.3279	0.2272	0.2006	0.2400
	100	0.1168	0.1112	0.1318	0.1109	0.1126	0.1189
	200	0.0708	0.0606	0.0639	0.0578	0.0607	0.0620
	500	0.0244	0.0250	0.0259	0.0239	0.0251	0.0253
	1000	0.0143	0.0141	0.0142	0.0133	0.0141	0.0142

Table 5: $\psi_5(t) = 1 - 48\,t + 218\,t^2 - 315\,t^3 + 145\,t^4$.

	n	BIC	AICc	HQ	S	T	PRESS
Gaussian Error							
snr=0.65	50	0.2635	0.3612	1.7660	0.4526	0.3779	0.6061
	100	0.1633	0.3104	0.6456	0.2429	0.3228	0.3661
	200	0.1020	0.1879	0.2475	0.1138	0.1888	0.2003
	500	0.0611	0.1056	0.1113	0.0668	0.1061	0.1095
	1000	0.0510	0.0811	0.0849	0.0547	0.0812	0.0831
snr=0.85	50	0.2000	0.2317	0.7109	0.2795	0.2374	0.3230
	100	0.0989	0.1441	0.2632	0.1148	0.1453	0.1699
	200	0.0681	0.1033	0.1303	0.0745	0.1039	0.1098
	500	0.0521	0.0632	0.0663	0.0540	0.0633	0.0643
	1000	0.0465	0.0503	0.0513	0.0459	0.0503	0.0504
snr=0.95	50	0.1527	0.1694	0.3760	0.1963	0.1718	0.2174
	100	0.0781	0.0953	0.1465	0.0845	0.0985	0.1077
	200	0.0570	0.0684	0.0771	0.0585	0.0685	0.0696
	500	0.0450	0.0442	0.0455	0.0419	0.0443	0.0446
	1000	0.0368	0.0334	0.0334	0.0339	0.0334	0.0334
Centered Gamma Error							
snr=0.65	50	0.2669	0.3244	1.8680	0.4057	0.3476	0.5814
	100	0.1617	0.3009	0.6513	0.2079	0.3297	0.3722
	200	0.0973	0.1885	0.2790	0.1116	0.1911	0.2058
	500	0.0612	0.1108	0.1280	0.0662	0.1124	0.1171
	1000	0.0513	0.0778	0.0829	0.0546	0.0781	0.0812
snr=0.85	50	0.1884	0.2197	0.7900	0.2650	0.2270	0.3229
	100	0.1048	0.1585	0.3120	0.1316	0.1741	0.2032
	200	0.0693	0.1229	0.1493	0.0765	0.1239	0.1265
	500	0.0524	0.0656	0.0696	0.0544	0.0658	0.0668
	1000	0.0470	0.0514	0.0522	0.0463	0.0515	0.0514
snr=0.95	50	0.1429	0.1576	0.3421	0.1767	0.1585	0.1851
	100	0.0748	0.0944	0.1358	0.0851	0.0951	0.1065
	200	0.0558	0.0695	0.0793	0.0569	0.0702	0.0733
	500	0.0444	0.0456	0.0462	0.0425	0.0457	0.0457
	1000	0.0373	0.0342	0.0342	0.0343	0.0342	0.0340
Uniform Error							
snr=0.65	50	0.2600	0.3515	1.7317	0.5167	0.3737	0.6430
	100	0.1579	0.2822	0.6103	0.2056	0.3015	0.3709
	200	0.1044	0.1875	0.2820	0.1149	0.1888	0.2081
	500	0.0602	0.0990	0.1118	0.0663	0.1003	0.1040
	1000	0.0509	0.0758	0.0780	0.0531	0.0759	0.0763
snr=0.85	50	0.1786	0.2301	0.7175	0.2484	0.2331	0.3087
	100	0.1005	0.1443	0.2767	0.1217	0.1513	0.1756
	200	0.0678	0.1068	0.1348	0.0724	0.1081	0.1157
	500	0.0514	0.0657	0.0716	0.0523	0.0660	0.0667
	1000	0.0463	0.0510	0.0514	0.0455	0.0510	0.0512
snr=0.95	50	0.1543	0.1725	0.3632	0.2001	0.1764	0.2178
	100	0.0790	0.0959	0.1454	0.0878	0.0964	0.1077
	200	0.0567	0.0692	0.0762	0.0581	0.0694	0.0716
	500	0.0444	0.0456	0.0467	0.0424	0.0456	0.0458
	1000	0.0373	0.0338	0.0338	0.0339	0.0338	0.0338

Table 6: $\psi_6(t) = 0.3\exp(-64(t-.25)^2) + 0.7\exp(-256(t-.75)^2)$.

	n	BIC	AICc	HQ	S	T	PRESS
Gaussian Error							
snr=0.65	50	0.7655	1.0836	5.0320	1.2000	1.1393	1.4901
	100	0.7020	1.1661	2.1737	0.8439	1.1947	1.3220
	200	0.6419	0.9985	1.3474	0.6253	1.0096	1.1309
	500	0.4290	0.5208	0.5500	0.3513	0.5253	0.5251
	1000	0.2467	0.3274	0.3438	0.2498	0.3320	0.3344
snr=0.85	50	0.6656	0.7496	2.3707	0.9529	0.7609	1.1440
	100	0.5353	0.5880	0.9913	0.5120	0.6157	0.6555
	200	0.3437	0.4287	0.4876	0.3287	0.4314	0.4412
	500	0.2139	0.2635	0.2750	0.2236	0.2636	0.2655
	1000	0.1879	0.1823	0.1852	0.1768	0.1823	0.1830
snr=0.95	50	0.4066	0.4205	0.8989	0.4842	0.4281	0.5326
	100	0.2700	0.3212	0.4777	0.2875	0.3248	0.3620
	200	0.2127	0.2297	0.2486	0.2155	0.2303	0.2363
	500	0.1655	0.1191	0.1200	0.1280	0.1191	0.1188
	1000	0.1077	0.0724	0.0724	0.0724	0.0724	0.0728
Centered Gamma Error							
snr=0.65	50	0.7908	1.0304	5.6731	1.2359	1.1515	1.6671
	100	0.6666	1.0793	2.4654	0.8090	1.1604	1.3408
	200	0.6130	0.8390	1.1636	0.6194	0.8421	0.8765
	500	0.4360	0.5266	0.5903	0.3743	0.5341	0.5476
	1000	0.2490	0.3357	0.3567	0.2418	0.3358	0.3399
snr=0.85	50	0.6737	0.7484	2.0863	0.8230	0.7552	0.9239
	100	0.5133	0.6437	1.1074	0.5813	0.6713	0.7006
	200	0.3592	0.4394	0.5298	0.3397	0.4519	0.4723
	500	0.2164	0.2632	0.2783	0.2259	0.2642	0.2655
	1000	0.1897	0.1772	0.1781	0.1813	0.1771	0.1770
snr=0.95	50	0.4247	0.4192	0.8852	0.4747	0.4183	0.5121
	100	0.2769	0.3197	0.4313	0.2980	0.3248	0.3444
	200	0.2122	0.2249	0.2523	0.2103	0.2250	0.2278
	500	0.1667	0.1222	0.1243	0.1342	0.1226	0.1218
	1000	0.1042	0.0728	0.0729	0.0736	0.0727	0.0727
Uniform Error							
snr=0.65	50	0.7189	1.1792	5.7416	1.4728	1.3589	1.9794
	100	0.6762	1.0807	2.1893	0.7817	1.0984	1.2609
	200	0.6208	0.8705	1.2099	0.6604	0.8729	0.9545
	500	0.4296	0.5109	0.5474	0.3322	0.5118	0.5154
	1000	0.2454	0.3321	0.3455	0.2397	0.3326	0.3379
snr=0.85	50	0.6838	0.7620	2.2574	0.8946	0.7563	1.0579
	100	0.5345	0.5978	0.9791	0.5490	0.6134	0.6909
	200	0.3468	0.4223	0.5317	0.3359	0.4262	0.4417
	500	0.2115	0.2814	0.2986	0.2226	0.2813	0.2864
	1000	0.1889	0.1772	0.1792	0.1812	0.1773	0.1766
snr=0.95	50	0.4234	0.4134	0.9229	0.4724	0.4174	0.5153
	100	0.2816	0.3218	0.4754	0.3027	0.3389	0.3634
	200	0.2104	0.2349	0.2577	0.2127	0.2378	0.2441
	500	0.1657	0.1247	0.1271	0.1297	0.1246	0.1254
	1000	0.1055	0.0773	0.0774	0.0771	0.0773	0.0771

for $n \geq 200$, the calculated $RMISE$ is not larger than 14% (it reaches 3% for $n = 1000$) when $snr = 0.85$ and 6% when $snr = 0.95$ (with $RMISE = 1.3\%$ for $n = 1000$). BIC gives the best results for small samples, even if $RMISE$ values are rather high.

For the models whose coefficient is ψ_5, simulations indicate Shibata's criterion as a good one for $n \geq 200$ whereas BIC is good for small samples. In fact when $n \geq 100$, $RMISE$ obtained using the Schwartz's method is not larger than 11% when $snr = 0.85$ or 0.95, and S performes similarly or better for $n \geq 200$. The HQ method is the worst one, even if all methods seem to be equivalents for $n = 1000$ and $snr = 0.95$.

When we estimate the coefficient ψ_6, we obtain very bad performances in general: when n and snr are large all selectors perform poorly with $RMISE$ not lower than $7 - 8\%$. For example, when $snr = 0.85$ and $n = 1000$, the best criterions give $RMISE$ around 18% (for example in the gaussian case the Shibata's criterion is the best with $RMISE = 17.7\%$).

At this point, in the light of previous comments, we can give some general practical indications. It seems that for large samples ($n \geq 500$) the Shibata's selector gives good results: in general, it supplies the best performances or at least near to the best ones, for any error distribution and signal-to-noise ratio.

When the simple size is small ($n \leq 200$), BIC is performing often better than the others one, even if the $RMISE$ is high. We may note that it gives better results than the corrected AIC, which should be the a priori suitable selector for small samples. The criterion that supplies the worst results is that of Hannan and Quinn: usually it performs very badly and in the best cases it gives $RMISE$ at least of the order of the other methods.

For the others selection methods, it seems that Rice's T, PRESS and AIC have a quite similar behavior for large samples, but in general they have not homogeneous and good performances in the different analyzed cases. We note also that the PRESS criterion, which is very used in the applications and in the software implementations, is not optimal also with large samples.

5 Conclusions and perspectives

As we have already noticed in this paper, when we use the Principal Component Functional Regression Estimator in order to estimate the functional linear regression model, we can introduce many techniques to choose the number of Principal Components: as simulations show, it seems that some criteria are better than others, and in particular the Shibata's criterion gives always good performances for large samples. These results remain valid also when models are not gaussian.

The attainment of convergence results in L^2-norm is a gap which has to be filled: if we had these results we could verify if, using the above mentioned selection criteria, the estimator attains the optimal L^2-rate of convergence.

As we previously mentioned, the PCR estimator which we have used here is not the only one which needs the choice of a regularization parameter. In particular the suggested results should be read in the case in which estimation is based on Fourier

Basis Expansion: in this case the functions of the base are chosen a priori and there is the need to identify their number.

When we use the Penalized spline method, the problem does not lie in the choice of the number of the knots, but in the smoothing parameter which controls the roughness of the curves (see Marx and Eilers, 1996, for a discussion on this topic).

Another perspective is the possibility of extending the results obtained in the i.i.d. context to the case of dependent observations: the same kind of analysis presented here should be made in the time series framework using functional linear models. This approach should lead to use this tool in a dependent context: also in this case we need some robust selection criterion to face the dimensionality problem.

References

[1] Akaike, H. (1973): Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki (Eds): *2nd International Symposium of Information Theory*, 267-81. Budapest: Akademia Kiado.

[2] Allen, D.M. (1974): The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, **16**, 125-127.

[3] Ash, R.B. and Gardner, M.F. (1975): *Topics in Stochastic Processes*. New York: Academic Press.

[4] Bosq, D. (1991): Modelisation, nonparametric estimation and prediction for continuous time series. In G. Roussas (Ed): *Nonparametric Functional Estimation and Related Topics*, 509-529. Kluwer Academic Press.

[5] Bosq, D. (2000): *Linear Processes in Function Spaces. Theory and Applications*. Lecture Notes in Statistics, **149**, Springer-Verlag.

[6] Burnham, K.P. and Anderson, D.R. (2002): *Model Selection and Multimodel Inference* II edition, Springer.

[7] Cardot, H., Ferraty, F.m and Sarda, P. (1999a): Functional linear model. *Statistic and Probability Letters*, **45**, 11-22.

[8] Cardot, H., Ferraty, F., and Sarda, P. (1999b): Spline estimators for the functional linear model: consistency, application and Splus implementation. *Rapport UBIA Toulouse*, **1999/1**.

[9] Cardot, H., Ferraty, F., and Sarda, P. (2002): Etude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. *Comptes Rendus de l'Académie des Sciences de Paris*, **330**, Série I, 501-504.

[10] Cardot, H., Ferraty, F., and Sarda, P. (2003): Spline estimators for the functional linear model. *Statistica Sinica*, to appear.

[11] Cavanaugh, J.E. (1997): Unifying the derivation for the Akaike and Corrected Akaike Information Criteria. *Statistics & Probability Letters*, **33**, 201-208.

[12] Dauxois, J., Pousse, A., and Romain, Y. (1982): Asymptotic theory for the principal component analysis of a random vector function: Some applications to statistical inference. *Journal of Multivariate Analysis*, **12**, 136-154.

[13] Ferraty, F. and Vieu, P. (2003): The functional nonparametric model and application to spectrometric data. *Computational Statistics and Data Analysis*, to appear.

[14] Hannan, E.J. and Quinn, B.G. (1979): The determination of order of autoregression. *Journal of the Royal Statistical Society B*, **41**, 190-195.

[15] Hastie, T. and Mallows, C. (1993): A discussion of "A statistical view of some chemometrics regression tools" by I.E. Frank. and J.H. Friedman. *Technometrics*, **53**, 140-143.

[16] Hurvich, C.M. and Tsai C.L. (1989): Regression and time series model selection in small samples. *Biometrika*, **76**, 297-307.

[17] McQuarrie, A.D.R. and Tsai C.L. (1998): *Regression and time series model selection*. World Scientific, Singapore.

[18] Marx, B.D. and Eilers P.H. (1996): Generalized linear regression on sampled signals with penalized lekelihood. In A. Forcina, G. M. Marchetti, R. Hatzinger, and G. Galmacci (Eds.): *Statistical Modelling. Proceeding of the 11th International workshop on statistical modelling*, Orvieto.

[19] Marx, B.D. and Eilers P.H. (1999): Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, **41**, 1-13.

[20] Miller, A.J. (1990): *Subset Selection in Regression*. Chapman and Hall.

[21] Ramsay, J.O. and Silverman, B.W. (1997): *Functional Data Analysis*. Springer-Verlag.

[22] Rice, J. (1984): Bandwidth choice for nonparametric regression. *The Annals of Statistics*, **12**, 1215-1230.

[23] Schwartz, G. (1978): Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.

[24] Shibata, R. (1981): An optimal selection of regression variables. *Biometrika*, **68**, 45-54.