

A note on functional linear regression*

Amparo Baíllo[†]

Universidad Autónoma de Madrid, 28049 Madrid, Spain

Abstract

This work focuses on the problem of linear regression with functional covariate and scalar response. We compare the performance of two linear (parametric) and one nonparametric (kernel) regression estimators via a Monte Carlo simulation study and the analysis of two real data sets. For one of the linear estimators, studied by Cai and Hall (2006), a new cut-off method is proposed, which improves over the thresholding procedure used by these authors.

Keywords: Functional linear regression; regularization; covariance eigenfunction; Fourier basis; cross-validation; roughness penalty.

AMS Classification 2000: 62J05

1 Introduction

We consider the linear regression problem with functional auxiliary variable X defined on an interval $\mathcal{T} = [0, T]$ and scalar response Y , that is, we assume that the following model holds

$$Y = a + \int_{\mathcal{T}} bX + \epsilon, \quad (1)$$

where $a \in \mathbb{R}$, $b \in L^2(\mathcal{T})$ and ϵ has zero mean and finite variance. Note that it is possible to express model (1) as $Y = a + \langle b, X \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the $L^2(\mathcal{T})$ inner product. From now on it is assumed that $a = 0$ and $\mathcal{T} = [0, 1]$.

We observe n random, independent copies of (X, Y) : (X_i, Y_i) , for $i = 1, \dots, n$. This sample is only recorded on an equispaced grid t_1, t_2, \dots, t_N of $[0, T]$ whose internodal space is $w = T/N$. Our aim is to estimate the regression function $m(x) := \int_{\mathcal{T}} bx$, using the information contained in the sample. This specific problem has been tackled, for instance, by Cardot, Ferraty and Sarda (2003), Ramsay and Silverman (2005) (Chapter 15) and Cai and Hall (2006).

As pointed out in Ramsay and Silverman (2005), minimization of the residual sum of squares

$$\text{SSE}(b) := \sum_{i=1}^n (Y_i - \langle X_i, b \rangle)^2$$

*Research partially supported by the IV PRICIT program titled *Modelización Matemática y Simulación Numérica en Ciencia y Tecnología* (SIMUMAT) and by Spanish grant MTM2004-00098.

[†]Address for correspondence: amparo.baillo@uam.es

yields a regression estimator which adapts perfectly to the sample points but is not very informative. Cai and Hall (2006) express this idea stating that this is an infinite-dimensional problem. This is why it is necessary an intermediate step of smoothing or regularization, which reduces the dimension of parameter b . A standard approach is to expand b and X_i using an orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $L^2(\mathcal{T})$,

$$b = \sum_{j=1}^{\infty} b_j \phi_j \quad \text{and} \quad X_i = \sum_{j=1}^{\infty} c_{ij} \phi_j \quad (2)$$

with $b_j = \langle b, \phi_j \rangle$ and $c_{ij} = \langle X_i, \phi_j \rangle$. The system $\{\phi_j\}_{j=1}^{\infty}$ can be, for example, the Fourier basis (see Ramsay y Silverman 2005) or the eigenfunctions of the covariance operator of X (see Cai y Hall 2006). Using the expansions in (2) and Parseval's identity, the residual sum of squares can be reexpressed as

$$\text{SSE}(b) = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^{\infty} b_j c_{ij} \right)^2.$$

The regularization step in this case reduces to truncating all the series at, say, the J -th term and using the approximations $b \simeq \sum_{j=1}^J b_j \phi_j$, $X_i \simeq \sum_{j=1}^J c_{ij} \phi_j$ and

$$\text{SSE}(b) \simeq \text{TSSE}(b) := \sum_{i=1}^n \left(Y_i - \sum_{j=1}^J b_j c_{ij} \right)^2. \quad (3)$$

The value of J , with $1 \leq J \leq n$, is called by Cai and Hall (2006) the frequency cut-off and, in a sense, it may be considered a resolution level.

In Section 2 we will specify in detail how to carry out the estimation of the “slope” b and the regression function m , first when $\{\phi_j\}_{j=1}^{\infty}$ is the Fourier basis and secondly when the series expansion is done in terms of the covariance eigenfunctions. The last part of Section 2 introduces the nonparametric kernel estimator of m . Our aim is to compare the performance of the three resulting regression estimators under the parametric linear model given in (1). This is done in Section 3 via a Monte Carlo study and the analysis of two real data sets. To our knowledge this is the first simulation study comparing the performance of the two parametric, linear regression estimators. The kernel estimator is included in the study in order to quantify the reduction of the error attained by incorporating all the available information on the model.

2 Estimation of the linear regression function

2.1 The linear estimator in terms of the Fourier basis

In this subsection we will use the (orthonormal) Fourier basis

$$\phi_1(t) = 1, \quad \phi_{2r}(t) = 2 \sin(2\pi r t), \quad \phi_{2r+1}(t) = 2 \cos(2\pi r t), \quad r = 1, 2, \dots$$

Let us denote by J_T the frequency cut-off used with this basis, that is, $\max(2r, 2r+1) \leq J_T$.

The first estimator of the parameter b that we will consider is the function $\hat{b} = \sum_{j=1}^{J_T} \hat{b}_j \phi_j$ minimizing $\text{TSSE}(\hat{b})$. This is equivalent to the system of equations $\mathbf{C}'\mathbf{C}\hat{\mathbf{b}} = \mathbf{C}'\mathbf{Y}$, where $\hat{\mathbf{b}} :=$

$(\hat{b}_1, \dots, \hat{b}_{J_T})'$, $\mathbf{Y} := (Y_1, \dots, Y_n)'$ and \mathbf{C} is the $n \times J_T$ matrix whose components are c_{ij} . Thus $\hat{\mathbf{b}} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{Y}$. Then, for $x \in L^2(\mathcal{T})$, the predicted value of Y at x is given by

$$\hat{m}_T(x) := \hat{\mathbf{b}}'\mathbf{c}(x) \quad (4)$$

where $\mathbf{c}(x) := (c_1(x), \dots, c_{J_T}(x))'$ and $c_j(x) := \langle x, \phi_j \rangle$.

There remains to determine the cut-off J_T in an automatic, data-driven way. As Ramsay and Silverman (2005) point out, in order to obtain reasonable results, the number J_T of terms in the Fourier expansion should be low. The reason is that J_T is the new dimension of the covariates after these have been expressed in terms of their Fourier expansion, so the lower is J_T the less we are affected by the curse of dimensionality. Here we have chosen J_T using the following cross-validation procedure

$$J_T := \arg \min_J \text{CVT}(J), \quad (5)$$

where $\text{CVT}(J) = \sum_{i=1}^n (Y_i - \hat{m}_{T,-i}(X_i))$ and $\hat{m}_{T,-i}$ is the linear estimator of m constructed in terms of the Fourier basis, as described in this subsection, but based on the sample

$$\mathcal{X}^{(-i)} := \{(X_1, Y_1), \dots, (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), \dots, (X_n, Y_n)\}.$$

The second way of estimating b is via the roughness penalty approach (see Chapter 10 in Ramsay and Silverman 2005). Here the aim is, not only to minimize the residual sum of squares, but also to prevent \hat{b} from fluctuating too much. This is done by minimizing the penalized residual sum of squares

$$\text{PSSE}_\lambda(b) := \sum_{i=1}^n (Y_i - \langle X_i, b \rangle)^2 + \lambda \int_{\mathcal{T}} [b''(s)]^2 ds = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^{J_T} b_j c_{ij} \right)^2 + \lambda \sum_{j=1}^{J_T} b_j^2 w_j^4,$$

where $\lambda > 0$ is a smoothing parameter and $\phi_j'' = -w_j^2 \phi_j$ for $j = 1, \dots, J_T$ (that is, $w_1 = 0$ and $w_j = 2\pi r$ for all $j \geq 2$). The L^2 norm of b'' summarizes the curvature in b and λ controls the trade-off between smoothness and adaptation to the data. In this approach the number of terms J_T in the Fourier expansion are no longer a key value as far as regularization is concerned, so in the Monte Carlo study we will take $J_T = 30$. Observe that the penalized sum of squared residuals can be expressed as

$$\text{PSSE}_\lambda(b) = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^{J_T} b_j c_{ij} \right)^2 + \lambda \mathbf{b}' \mathbf{K} \mathbf{b}, \quad (6)$$

where $\mathbf{b} := (b_1, \dots, b_{J_T})'$ and \mathbf{K} is the matrix with components $K_{jk} := \langle \phi_j'', \phi_k'' \rangle$. The value of \mathbf{b} minimising (6) is given by $\hat{\mathbf{b}}_{\text{pen}} = (\mathbf{C}'\mathbf{C} + \lambda \mathbf{K})^{-1}\mathbf{C}'\mathbf{Y}$ and the corresponding regression estimator is $\hat{m}_{T,\text{pen}}(x) = \hat{\mathbf{b}}_{\text{pen}}'\mathbf{c}(x)$.

In this work the smoothing parameter λ has been chosen in an automatic way as the one minimizing the cross validation score

$$\text{CVP}(\lambda) := \sum_{i=1}^n \left(Y_i - \sum_{j=1}^{J_T} \mathbf{b}_{\lambda,j}^{(-i)} c_{ij} \right)^2, \quad (7)$$

where $\mathbf{b}_{\lambda}^{(-i)} = \left(\mathbf{b}_{\lambda,1}^{(-i)}, \dots, \mathbf{b}_{\lambda,J_T}^{(-i)} \right)'$ minimizes $\text{PSSE}_\lambda(b)$ based on the sample $\mathcal{X}^{(-i)}$.

2.2 The linear estimator in terms of the covariance eigenfunctions

The approach followed by Cai and Hall (2006) to estimate b is based on functional principal components. Let us briefly describe the procedure. We assume that the covariance function of X , $V(s, t) := \text{cov}(X(s), X(t))$, is positive definite. Then V admits a spectral decomposition $V(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t)$, for $s, t \in \mathcal{T}$, where λ_j and ϕ_j denote respectively an eigenvalue and its corresponding eigenfunction for the linear operator with kernel V . Let the eigenvalues be arranged in decreasing order and observe that the eigenfunctions $\{\phi_j\}_{j=1}^{\infty}$ form an orthonormal basis in $L^2(\mathcal{T})$.

Since the operator V is unknown, we approximate it by its empirical analogue

$$\hat{V}(s, t) := \frac{1}{n} \sum_{i=1}^n (X_i(s) - \bar{X}(s))(X_i(t) - \bar{X}(t)) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(s) \hat{\phi}_j(t),$$

where $\bar{X} := n^{-1} \sum_{i=1}^n X_i$. Just as in the case of V , $\hat{\lambda}_j$ denotes an eigenvalue for the linear operator with kernel \hat{V} , $\hat{\phi}_j$ is the corresponding eigenfunction and the eigenvalues are ordered ($\hat{\lambda}_1 > \hat{\lambda}_2 > \dots$). Observe that $\hat{\lambda}_j = 0$ for $j \geq n + 1$.

As the pair (λ_j, ϕ_j) is approximated by $(\hat{\lambda}_j, \hat{\phi}_j)$, the coefficient $b_j = \langle b, \phi_j \rangle$ will be estimated in the following way. First observe that $b_j = g_j / \lambda_j$, where $g_j = \langle g, \phi_j \rangle$ and $g(t) = \langle V(t, \cdot), b \rangle$. A consistent estimator of g is given by

$$\hat{g}(t) = \frac{1}{n} \sum_{i=1}^n [X_i(t) - \bar{X}(t)](Y_i - \bar{Y})$$

with $\bar{Y} := n^{-1} \sum_{i=1}^n Y_i$. Thus we will estimate b_j by $\hat{b}_j = \hat{g}_j / \hat{\lambda}_j$, where $\hat{g}_j = \langle \hat{g}, \hat{\phi}_j \rangle$.

In this case the cut-off will be denoted by J_E . Once this value has been fixed, the regression estimator evaluated at $x \in L^2(\mathcal{T})$ is given by

$$\hat{m}_E(x) := \sum_{j=1}^{J_E} \hat{b}_j \hat{c}_j(x), \quad (8)$$

where now $c_j(x) = \langle x, \phi_j \rangle$.

Cai and Hall (2006) suggest the following thresholding algorithm to determine J_E : let $C > 0$ and $0 < c \leq 1/2$, then choose $J_E^{TH} := \max\{j \geq 1 : \hat{\lambda}_j \geq t := Cn^{-c}\}$. An appropriate choice when the function X is observed on a regular grid seems to be $c = 1/2$, so we will take this value. Regarding the constant C , we have considered different values ($C = 0.01, 0.1, \dots, 0.2$), as in Cai and Hall (2006). However, as the simulations in Section 3 reveal, this choice of the cut-off yields a very large number of terms in the eigenfunction expansion. Thus, the squared prediction error attained by the corresponding regression estimator is too large even for a fixed x , and it displays an erratic behaviour (not converging to 0) when x is chosen at random. This is why in this work the cut-off J_E at which the series is truncated will be determined by a cross-validation procedure similar to the one proposed in Subsection 2.1

$$J_E^{CV} := \arg \min_J \text{CVE}(J), \quad (9)$$

where $\text{CVE}(J) = \sum_{i=1}^n (Y_i - \hat{m}_{E,-i}(X_i))^2$ and $\hat{m}_{E,-i}$ is the linear estimator of m constructed in terms of the eigenfunction basis, as explained in this subsection, and based on the sample $\mathcal{X}^{(-i)}$.

2.3 The kernel estimator

In this subsection we consider a nonparametric estimator of the regression function, concretely the functional kernel estimator given by

$$\hat{m}_K(x) := \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)}, \quad (10)$$

where K is an asymmetrical decreasing kernel function, $K_h(\cdot) := K(\|\cdot\|/h)$ and $h = h_n$ is a positive smoothing parameter. This regression estimator has been studied, for instance, by Ferraty and Vieu (2006). Here we will use the asymmetrical Gaussian kernel $K(t) = \sqrt{2/\pi} \exp(-t^2/2)$ for $t \in (0, \infty)$.

The bandwidth has been chosen via the following cross-validation procedure described in Rachdi and Vieu (2005) (see also Ferraty and Vieu 2006, p. 101)

$$h_{\text{opt}} = \arg \min_h \text{CVK}(h), \quad (11)$$

where $\text{CVK}(h) = \sum_{i=1}^n (Y_i - \hat{m}_{K,-i}(X_i))^2$ and $\hat{m}_{K,-i}$ is the kernel estimator of m based on the observations from $\mathcal{X}^{(-i)}$.

3 Comparison of the functional regression estimators

In this section we will compare the performance of the regression estimators described in Section 2, first via a Monte Carlo simulation study (see Subsection 3.1) and in Subsection 3.2 via the analysis of real climate data from U.S.

3.1 The Monte Carlo study

The linear models generating the simulated data have been used in Cai and Hall (2006) and Hall and Horowitz (2004). The dependent variable Y in all the models is assumed to be given by (1), with $a = 0$ and $\mathcal{T} = [0, 1]$. Initially, in Subsection 3.1.1, we consider in more detail what we call Model 1, taken from Cai and Hall (2006), and we describe the problem encountered with the cut-off procedure proposed by these authors. In Subsection 3.1.2 we will introduce the rest of the models and carry out the main simulation study.

More specifically, in Model 1 the random functions X_i are independent and identically distributed as

$$X(t) = \sum_{j=1}^{50} Z_j 2^{1/2} \cos(j\pi t),$$

where Z_1, \dots, Z_{50} are independent and Z_j follows a normal distribution with mean 0 and variance $4j^{-2}$, for $j = 1, \dots, 50$. The “slope” of the linear model is given by

$$b(t) = \sum_{j=1}^{50} j^{-4} 2^{1/2} \cos(j\pi t)$$

and the error ϵ is normally distributed with mean 0 and variance 4.

3.1.1 Squared prediction error for fixed x

Here we consider the case where prediction of Y in Model 1 is aimed only at a fixed value of X , $x = \sum_{j=1}^{50} j^{-2} 2^{1/2} \cos(j\pi t)$ (see Cai and Hall 2006). Then $m(x) = \sum_{j=1}^{50} j^{-6} = 1.0173$. In each of $B = 1000$ Monte Carlo samples the following experiment has been performed. We have taken a sample (X_i, Y_i) , $i = 1, \dots, n$, of size $n = 100$. Each X_i was observed discretely on an equally-spaced grid of $N = 200$ points.

In order to point out the problems encountered with the cut-off selection procedure proposed in Cai and Hall (2006), the squared prediction error $\text{SE}(x) := (\hat{m}(x) - m(x))^2$ has been computed for $\hat{m} = \hat{m}_T$ with $J_T = 3$, $J_T = 4$ and $J_T = 5$ and for $\hat{m} = \hat{m}_E$ with $C = 0.01, 0.1, 1, 2, 3$. Table 1 displays the average and quartiles of the observed $\text{SE}(x)$ for both regression estimators, computed over the $B = 1000$ simulations, as well as the average value of J_E^{TH} for the different values of C .

From the results in Table 1 it is clear that the estimator \hat{m}_T constructed in terms of the Fourier series performs much better than \hat{m}_E , the one constructed in terms of the covariance eigenfunctions, at least for the values of C considered in Cai and Hall (2006). It is also clear that, in the case of \hat{m}_E , the squared prediction error decreases as C increases. As a matter of fact, in Table 2 we can see the same statistics of the squared prediction error attained by \hat{m}_E when J_E is fixed and takes the values $1, \dots, 6$. Observe that the error is lower than in the thresholding procedure and seems to reach a minimum around $J = 1$ or 2 . This suggests that, in the cross-validation procedure for selecting J_E , the range over which the minimum appearing in (9) is computed could be $1 \leq J \leq 10$. We have restricted to this range for the rest of this work. Similarly, from now on the minimum in (5) will be computed for $1 \leq J \leq 15$.

Table 3 displays the average and quartiles of the squared prediction error $E(x)$ for \hat{m}_T , \hat{m}_E , $\hat{m}_{T,\text{pen}}$ and \hat{m}_K computed over $B = 1000$ simulations. The number of terms, J_T and J_E^{CV} , in the series expansions of the first two estimators are chosen as in (5) and (9) respectively. In Table 3 we can also see the average value of J_T and J_E^{CV} , which is coherent with the results of Tables 1 and 2. The window width in the kernel estimator is determined through the cross-validation procedure in (11). The error attained by the linear regression function expanded in terms of the covariance eigenfunction basis is now smaller than the one obtained with the Fourier expansion.

3.1.2 Squared prediction error for random x

In this subsection we consider a better measure of the performance of the regression estimators, $\text{SE}(X_b)$, for $b = 1, \dots, B$ simulations, where X_b is randomly generated from the distribution of the auxiliary variable X in the linear model. Apart from Model 1, we consider two more models. Model 2 was also introduced in Cai and Hall (2006) and is the same as Model 1 except for the parameter b , now given by $b(t) = 10 \sum_{j=1}^{50} j^{-2} 2^{1/2} \cos(j\pi t)$.

In Model 3, used in Hall and Horowitz (2004), the “slope” is given by $b = \sum_{j=1}^{50} b_j \phi_j$, with $\phi_1 \equiv 1$, $\phi_{j+1} = 2^{1/2} \cos(j\pi t)$ for $j \geq 1$, $b_1 = 0.3$ and $b_j = 4(-1)^{j+1} j^{-2}$ for $j > 1$. The auxiliary variable is generated as $X(t) = \sum_{j=1}^{50} \gamma_j Z_j \phi_j$, where the Z_j ’s are independent and uniformly distributed on $[-3^{1/2}, 3^{1/2}]$, $\gamma_j = (-1)^{j+1} j^{-\alpha/2}$ for $\alpha = 1.1$ and 2 . The error ϵ follows a normal $N(0, \sigma_\epsilon^2)$ distribution, with $\sigma_\epsilon = 0.5$ and 1 .

Table 4 displays the average, median and quartiles of the sample $\{\text{SE}(X_b) : b = 1, \dots, B\}$, computed over $B = 1000$ simulations, for the estimators \hat{m}_T , \hat{m}_E , $\hat{m}_{T,\text{pen}}$ and \hat{m}_K and for Models 1, 2 and 3. The cut-offs J_T and J_E , the penalization parameter λ and the window width h are chosen via the cross-validation procedures specified in (5), (9), (7) and (11) respectively. The auxiliary variable X was observed on grids of $N = 50$ and $N = 200$ equi-spaced points.

The results appearing in Table 4 indicate that increasing the number of grid points from $N = 50$ to $N = 200$ does not necessarily reduce the prediction error, especially as far as the linear regression estimator is concerned. We can see that in *all* cases the linear estimator based on the eigenfunction expansion performs better than any of the other estimators: the mean and the median prediction error is the smallest. This suggests that, if the linear model assumption holds, then the parametric estimator proposed by Cai and Hall (2006), with the cut-off J_E selected as in (9), will be a better choice than the linear regression estimator based on the Fourier expansion.

Regarding the linear estimators \hat{m}_T and $\hat{m}_{T,\text{pen}}$, in general the latter performs better than the former. In fact, in Model 3, for $\alpha = 2$ and $\sigma_\epsilon = 1$, the kernel estimator nearly outperforms \hat{m}_T . This superiority of $\hat{m}_{T,\text{pen}}$ over \hat{m}_T will be particularly apparent in the analysis with real data developed in Subsection 3.2. Observe finally that, due to the fact that the underlying regression function is linear, the kernel regression estimator gives always the worst results. However, when analyzing real data, we will see that the nonparametric estimator is a reasonable and competitive choice. Further, comparing the results given by the kernel estimator with those attained by the parametric ones may give some insight into the adequacy of the linearity assumption (1) for the particular data of interest.

3.2 Climate data from U.S.

The aim of this subsection is to compare the performance of the linear regression estimators \hat{m}_T , \hat{m}_E and $\hat{m}_{T,\text{pen}}$ and the nonparametric one, \hat{m}_K , in the analysis of two real U.S.A. climate data sets. Both of them have been obtained from the U.S. National Climatic Data Center website (www.ncdc.noaa.gov).

In the first group of data we wish study to what extent the total number of tornados in each U.S. state along the period 2000-2005, can be predicted by the temperatures registered in the same locations along the same years. This is of interest, for instance, when assessing the possible consequences (like an increase in the number of extreme climatic events) of an overall increase in the temperatures due to the climatic change. The auxiliary variable X_i is the monthly average temperature (measured in °F) in state i from the year 2000 to 2005, for $i = 1, \dots, 48$ states, and the response variable Y_i is the square root of the total number of tornados observed in that state for the same period. In Figure 1 we can see the evolution of the temperature curves.

We assume that model (1) holds and we wish to compare the performance of the regression estimators \hat{m}_T , \hat{m}_E , $\hat{m}_{T,\text{pen}}$ and \hat{m}_K . In Table 5 we have approximated (via a cross-validation procedure) the squared prediction error attained by each of these estimators. Observe that the linear regression estimator computed with the eigenfunction procedure introduced by Cai and Hall (2006) is the one performing best. Further, from the fact that the medians of the errors

attained by both \hat{m}_E and $\hat{m}_{T,\text{pen}}$ are the smallest we may guess that the linear model yields a reasonable description of the regression function.

We now consider the second data set, where the covariate functions are the daily maximum temperatures (in °F) recorded in $n = 80$ weather stations from South Dakota in year 2000 (see Figure 2). The response variable Y is the logarithm of the total precipitation in each of the stations during the same year. In Table 6 we report the approximation to the mean and quartiles of the squared prediction error attained by the four regression estimators. Observe that, in this case, the kernel estimator \hat{m}_K is the one performing best, while the linear regression estimator based on the covariance eigenfunctions, \hat{m}_E , is not satisfactory at all. We also note that the value of J_T , selected via the cross-validation procedure given by (5), was extremely dependent on the range of values of J over which the minimum appearing in (5) was computed. The larger was this range, the larger would be J_T and the worse would perform \hat{m}_T . Other transformations of the precipitation data, like the square root, yield no significantly different results for any of the regression estimators. This suggests that the linear model is not an adequate description of the relationship between precipitation totals and temperatures in South Dakota along 2000. We note, however, that the linear estimator $\hat{m}_{T,\text{pen}}$ appears to be quite robust to misspecification of the model. Thus, under the assumption of a functional linear regression model like (1), when estimating the regression function m , it seems reasonable that at least two parametric estimators be used and their results compared before choosing one particular regression methodology.

4 Conclusions

In this work we have considered the linear regression model for functional auxiliary variable X and scalar response Y . We have compared the performance of three regression estimators under this model via a Monte Carlo study and also via the analysis of two real datasets. Two of the estimators considered are parametric, linear, and are characterized by the regularization of X in terms of the trigonometric basis and the covariance eigenfunctions respectively. The choice of the cut-off in these expansions has been addressed and, in the case of the second linear estimator, we have proposed a cut-off selection procedure which improves over previous existent ones. In future work it would be interesting to consider a roughness penalty approach for this linear estimator based on the covariance eigenfunctions expansion. The third regression estimator included in the analysis has been a nonparametric, kernel one.

The Monte Carlo study shows that, under the hypothesis of linearity, the parametric estimator based on the eigenfunction expansion outperforms the others. However, the examples with real data suggest that this estimator is highly dependent on the linearity assumption and that the parametric estimator based on trigonometric expansion with roughness penalty is more robust to misspecification of the model.

Acknowledgement. The author is grateful to Prof. Antonio Cuevas for careful reading of the manuscript and interesting comments.

References

- Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.*, 34, 2159-2179.
- Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 13, 571–591.
- Ferraty, F. and Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *C. R. Acad. Sci. Paris*, 330, Série I, 139142.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York.
- Hall, P. and Horowitz, J. L. (2004). Methodology and convergence rates for functional linear regression. Manuscript.
- Rachdi, M. and Vieu, P. (2005). Sélection automatique du paramètre de lissage pour l'estimation non paramétrique de la régression pour des données fonctionnelles. *C. R. Math. Acad. Sci. Paris*, 341, 365–368.
- Ramsay, J. O. and Silverman, B. (2005). *Functional Data Analysis*. Second edition. Springer-Verlag, New York.

Table 1: Squared prediction error for \hat{m}_T , \hat{m}_E and \hat{m}_K and fixed x over $B = 1000$ simulations

		\hat{m}_T			\hat{m}_E				
Error	Average	0.0152	0.0160	0.0166	0.0335	0.0322	0.0309	0.0263	0.0256
	1st quartile	0.0014	0.0013	0.0018	0.0034	0.0029	0.0033	0.0027	0.0022
	Median	0.0067	0.0070	0.0078	0.0144	0.0152	0.0132	0.0121	0.0108
	3rd quartile	0.0196	0.0202	0.0220	0.0416	0.0412	0.0403	0.0356	0.0327
		$J_T = 3$	$J_T = 4$	$J_T = 5$	$C = 0.01$	$C = 0.1$	$C = 1$	$C = 2$	$C = 3$
		Average J_E^{TH}			50	50	48	43	39

Table 2: Squared prediction error for \hat{m}_E and fixed x over $B = 1000$ simulations

Error	Average	0.0114	0.0124	0.0142	0.0151	0.0151	0.0155
	1st quartile	0.0012	0.0011	0.0015	0.0015	0.0013	0.0016
	Median	0.0052	0.0054	0.0061	0.0069	0.0066	0.0073
	3rd quartile	0.0146	0.0165	0.0186	0.0194	0.0207	0.0209
J_E		1	2	3	4	5	6

Table 3: Squared prediction error for \hat{m}_T and \hat{m}_E and fixed x , with cross-validation choice of J

		\hat{m}_T	\hat{m}_E	$\hat{m}_{T,\text{pen}}$	\hat{m}_K
Error	Average	0.0145	0.0132	0.0147	0.1524
	1st quartile	0.0014	0.0013	0.0016	0.0144
	Median	0.0069	0.0058	0.0069	0.0723
	3rd quartile	0.0191	0.0181	0.0204	0.2120
	Average J	3	2		

Table 4: Squared prediction error for \hat{m}_T , \hat{m}_E and \hat{m}_K and random x , over $B = 1000$ simulations.

		$N = 50$				$N = 200$			
		\hat{m}_T	\hat{m}_E	$\hat{m}_{T,\text{pen}}$	\hat{m}_K	\hat{m}_T	\hat{m}_E	$\hat{m}_{T,\text{pen}}$	\hat{m}_K
Error Model 1	Average	0.2413	0.1388	0.2928	0.4779	0.2476	0.1534	0.2444	0.4960
	1st quartile	0.0119	0.0055	0.0169	0.0299	0.0159	0.0038	0.0155	0.0038
	Median	0.0628	0.0294	0.0785	0.1403	0.0782	0.0261	0.0814	0.0261
	3rd quartile	0.2222	0.1127	0.2566	0.4704	0.2406	0.1106	0.2511	0.1106
	Average J	3	2			3	2		
Error Model 2	Average	0.6749	0.4378	0.5604	14.3632	0.6263	0.3264	0.5408	12.0954
	1st quartile	0.0644	0.0354	0.0523	0.7476	0.0520	0.0248	0.0492	0.6218
	Median	0.2771	0.1765	0.2096	3.4657	0.2852	0.1184	0.2406	3.3564
	3rd quartile	0.8322	0.5326	0.6626	10.2790	0.8111	0.3631	0.6234	10.2985
	Average J	9	6			8	5		
Error Model 3 $\alpha = 1.1$ $\sigma_\epsilon = 0.5$	Average	0.0293	0.0155	0.0187	0.0451	0.0326	0.0173	0.0228	0.0522
	1st quartile	0.0031	0.0014	0.0017	0.0040	0.0029	0.0013	0.0022	0.0054
	Median	0.0134	0.0060	0.0085	0.0202	0.0139	0.0066	0.0100	0.0248
	3rd quartile	0.0370	0.0192	0.0239	0.0573	0.0421	0.0221	0.0287	0.0725
	Average J	6	4			6	4		
Error Model 3 $\alpha = 1.1$ $\sigma_\epsilon = 1$	Average	0.0889	0.0586	0.0645	0.0880	0.0912	0.0616	0.0686	0.0881
	1st quartile	0.0083	0.0042	0.0048	0.0078	0.0089	0.0052	0.0065	0.0080
	Median	0.0345	0.0179	0.0227	0.0371	0.0412	0.0223	0.0268	0.0421
	3rd quartile	0.1032	0.0606	0.0727	0.1102	0.1157	0.0696	0.0803	0.1217
	Average J	4	4			4	4		
Error Model 3 $\alpha = 2$ $\sigma_\epsilon = 0.5$	Average	0.0209	0.0136	0.0139	0.0286	0.0227	0.0153	0.0168	0.0305
	1st quartile	0.0020	0.0009	0.0010	0.0025	0.0020	0.0012	0.0012	0.0028
	Median	0.0078	0.0048	0.0051	0.0130	0.0092	0.0056	0.0059	0.0132
	3rd quartile	0.0241	0.0155	0.0166	0.0360	0.0267	0.0165	0.0180	0.0387
	Average J	4	3			4	3		
Error Model 3 $\alpha = 2$ $\sigma_\epsilon = 1$	Average	0.0623	0.0501	0.0524	0.0640	0.0665	0.0536	0.0492	0.0676
	1st quartile	0.0047	0.0031	0.0036	0.0059	0.0050	0.0034	0.0037	0.0056
	Median	0.0242	0.0145	0.0176	0.0275	0.0239	0.0167	0.0174	0.0273
	3rd quartile	0.0729	0.0511	0.0577	0.0843	0.0834	0.0548	0.0567	0.0858
	Average J	3	3			3	3		

Table 5: Squared prediction error in regression estimation with tornado-temperature data

		\hat{m}_T	\hat{m}_E	$\hat{m}_{T,\text{pen}}$	\hat{m}_K
Error	Average	6.4146	2.8436	8.6807	4.5771
	1st quartile	0.3738	0.1680	0.2111	0.2082
	Median	2.4707	1.0185	0.6294	1.1768
	3rd quartile	6.1571	2.7823	2.6148	4.1547
Average J		12	10		

Table 6: Squared prediction error in regression estimation with precipitation-temperature data

		\hat{m}_T	\hat{m}_E	$\hat{m}_{T,\text{pen}}$	\hat{m}_K
Error	Average	0.0588	10.3013	0.0479	0.0354
	1st quartile	0.0096	9.3765	0.0083	0.0022
	Median	0.0368	10.4622	0.0233	0.0142
	3rd quartile	0.0840	11.0599	0.0617	0.0479
	Average J	16	1		

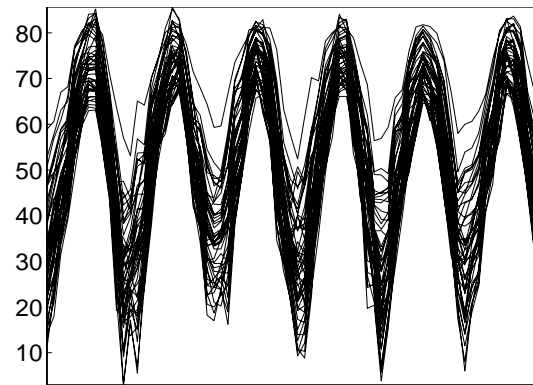


Figure 1: Average monthly temperatures in U.S.A. states from 2000 to 2005.

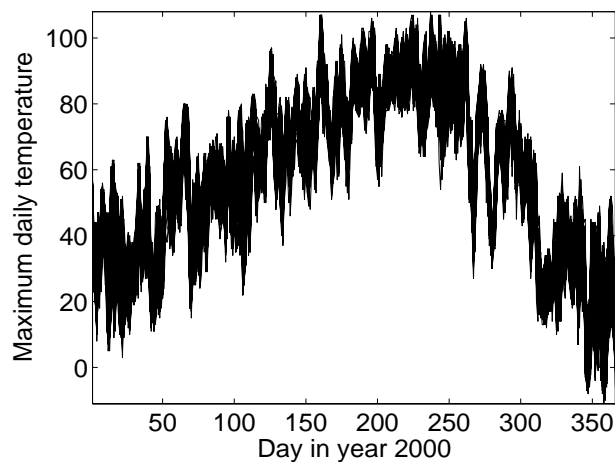


Figure 2: Daily maximum temperatures along year 2000 in 80 stations from South Dakota.