Local Linear Functional Regression based on Weighted Distance-Based Regression

Eva Boj [a] Pedro Delicado [b,*] Josep Fortiana [c]

[a] *Departament de Matemàtica Econòmica, Financera i Actuarial, Universitat de Barcelona, Barcelona, Spain.*

[b] *Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, Spain*

[c] *Departament de Probabilitat, Lògica i Estadística, Universitat de Barcelona, Barcelona, Spain.*

Abstract

We consider the problem of nonparametrically predicting a scalar response variable y from a functional predictor χ. We have n observations (χ_i, y_i) and we assign a weight $w_i \propto K\left(d(\chi, \chi_i)/h\right)$ to each χ_i, where $d(\,\cdot\,,\,\cdot\,)$ is a semi-metric, K is a kernel function and h is the bandwidth. Then we fit a Weighted (Linear) Distance-Based Regression, where the weights are as above and the distances are given by a possibly different semi-metric. This approach can be extended to nonparametric predictions from other kind of explanatory variables (e.g., data of mixed type) in a natural way.

Key words: Distance-based prediction, functional data analysis, local linear regression, nonparametric regression, weighted regression.

1 Introduction

Observing and saving complete functions as results of random experiments is nowadays possible by the development of real-time measurement instruments and data storage resources. For instance, continuous-time clinical monitoring is a common practice today. Functional Data Analysis (FDA) deals with the

* Corresponding author. *Phone:* (+34) 934015698. *Fax:* (+34) 934015855. *Postal address:* Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Edifici C5-214. C/ Jordi Girona 1-3, 08034, Barcelona, Spain. *E-mail address:* pedro.delicado@upc.edu

statistical description and modelization of samples of random functions. Functional versions for a wide range of statistical tools (ranging from exploratory and descriptive data analysis to linear models to multivariate techniques) have been recently developed. See Ramsay and Silverman [1] for a general perspective on FDA and Ferraty and Vieu [2] for a nonparametric approach. Special issues recently dedicated to this topic by several journals ([3], [4], [5]) bear witness to the interest on this topic in the Statistics community.

In this paper we consider the problem of predicting a scalar response using a functional predictor. Let us give an example. *Spectrometric Data* are described in Chapter 2 of [2]. This dataset includes information about 215 samples of chopped meat. For each of them, the function χ, relating absorbance versus wavelength, has been recorded for 100 values of wavelength in the range 850-1050 nm. An additional response variable is observed: y, the sample fat content obtained by analytical chemical processing. Given that obtaining a spectrometric curve is less expensive that determining the fat content by chemical analysis, it is important to predict the fat content y from the spectrometric curve χ. In Section 3.1 the Spectrometric Data are used to illustrate the methods we propose in this work.

In technical terms, the problem is stated as follows. Let $(\boldsymbol{\chi}, Y)$ be a random element where the first component $\boldsymbol{\chi}$ is a random element of a functional space (typically a real function $\boldsymbol{\chi}$ from $[a, b] \subseteq \mathbb{R}$ to $\mathbb{R}$) and Y is a real random variable. We consider the problem of predicting the scalar response variable y from the functional predictor χ. We assume that we are given n i.i.d. observations $(\chi_i, y_i), i = 1, \ldots, n$, from $(\boldsymbol{\chi}, Y)$ as a training set. Let $m(\chi) = E(Y|\boldsymbol{\chi} = \chi)$ be the regression function. Then an estimate of $m(\chi)$ is a good prediction of y. The linear functional regression model, considered in [1], assumes that

$$m(\chi) = \alpha + \int_a^b \chi(t)\beta(t)dt, \text{ and } y_i = m(\chi_i) + \varepsilon_i,$$

ε_i having zero expectation. The parameter β is a function and $\alpha \in \mathbb{R}$. The authors propose to estimate β and α by penalized least squares:

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \left(y_i - \alpha - \int_T \chi_i(t)\beta(t)dt \right)^2 + \lambda \int_a^b (L(\beta)(t))^2 dt,$$

where $L(\beta)$ is a linear differential operator giving a penalty to avoid too much rough β functions and $\lambda > 0$ acts as a smoothing parameter.

Ferraty and Vieu [2] consider this linear regression as a parametric model because only a finite number of functional elements is required to describe it (in this case only one is needed: β). They consider a nonparametric functional regression model where few regularity assumptions are made on the regression

function $m(\chi)$. They propose the following kernel estimator for $m(\chi)$:

$$\hat{m}_K(\chi) = \frac{\sum_{i=1}^n K(d(\chi,\chi_i)/h) y_i}{\sum_{i=1}^n K(d(\chi,\chi_i)/h)} = \sum_{i=1}^n w_i(\chi) y_i,$$

where $w_i(\chi) = K(d(\chi,\chi_i)/h)/\sum_{j=1}^n K(d(\chi,\chi_j)/h)$, K is a kernel function with support $[0,1]$, the bandwidth h is the smoothing parameter (depending on n), and $d(\cdot,\cdot)$ is a semi-metric ($d(\chi,\chi) = 0$, $d(\chi,\gamma) = d(\gamma,\chi)$, $d(\chi,\gamma) \leq d(\chi,\psi) + d(\psi,\gamma)$) in the functional space $\mathcal{F} = \{\chi : [a,b] \to \mathbb{R}\}$ to which the data χ_i belong. Examples of semi-metrics in $\mathcal{F}$ are L_2 distances between derivatives,

$$d_r^{deriv}(\chi,\gamma) = \left(\int_a^b \left(\chi^{(r)}(t) - \gamma^{(r)}(t) \right)^2 dt \right)^{1/2};$$

and the L_2 distance in the space of the first q functional principal components of the functional data set $\chi_i, i = 1,\ldots,n$: $d_q^{PCA}(\chi,\gamma) = (\sum_{k=1}^q (\psi_k^\chi - \psi_k^\gamma)^2)^{1/2}$, where ψ_k^χ is the score of the function χ in the k-th principal component. See Chapters 8 and 9 in [1] or Chapter 3 in [2] for more information about functional principal component analysis.

In [2] it is proved that $\hat{m}_K(\chi)$ is a consistent estimator (in the sense of almost complete convergence) of $m(\chi)$ under regularity conditions on m, $\boldsymbol{\chi}$ (involving small balls probability), Y and K. Moreover, Ferraty et al. [6] prove the mean squared convergence and asymptotic distribution of $\hat{m}_K(\chi)$.

The book of Ferraty and Vieu [2] lists several interesting open problems concerning nonparametric functional regression. In particular, their *Open Question 5* addresses the transfer of local polynomial regression ideas to an infinite dimensional setting in order to extend the estimator $\hat{m}_K(\chi)$, that is a kind of Nadaraya-Watson regression estimator.

A first answer to this question is given in Baíllo and Grané [7]. They propose a natural extension of the finite dimensional local linear regression, by solving the problem

$$\min_{\alpha,\beta} \sum_{i=1}^n w_i(\chi) \left(y_i - \alpha - \int_T (\chi_i(t) - \chi(t))\beta(t)dt \right)^2,$$

where local weights $w_i(\chi) = K(||\chi - \chi_i||/h)/\sum_{j=1}^n K(||\chi - \chi_j||/h)$ are defined by means of L_2 distances (it is assumed that all the functions are in $L_2([a,b])$). Their estimator of $m(\chi)$ is $\hat{m}_{LL}(\chi) = \hat{\alpha}$. Closely related approaches can be seen in [8] and [9].

In this work we give an alternative response to the same open question. Our proposal rests on Distance-Based Regression (DBR), a prediction tool based on inter-individual distances including Ordinary Least Squares Regression (OLS)

as a particular case (see Section 2). Specifically, we use Weighted Distance-Based Regression, the weighted version of DBR, where each case (χ_i, y_i) has a weight $w_i \propto K\left(d(\chi, \chi_i)/h\right)$. Subsection 2.2 presents all the formulas needed to implement the Weighted DBR. We name our proposal Local Linear Distance-Based Regression, and Section 3 is devoted to introduce it with detail, including the analysis of Spectrometric Data. Section 4 contains some concluding remarks.

2 Weighted Distance-Based Regression: Definition and results

Distance-Based Regression was introduced by Cuadras [10] in 1989 and has been developed in [11], [12] and [13]. Let $\Omega = \{\mathcal{O}_1, \ldots, \mathcal{O}_n\}$ be a set of n objects (or individuals or cases) randomly drawn from a population. For individual $\mathcal{O}_i$ we have observed the value y_i of a continuous one-dimensional response variable. We assume that a distance function δ (being a metric or semi-metric) is defined between the elements of Ω, usually based on predictors $\boldsymbol{Z}$ observed for every $\mathcal{O}_i \in \Omega$ as $\boldsymbol{z}_i$. Let $\boldsymbol{\Delta} = (d^2_{i,j})_{i=1..n, j=1..n}$ be the inter-individual squared distances matrix. The available information $\boldsymbol{Z}$ for the elements of Ω can be a mixture of quantitative and qualitative variables or, possibly, other nonstandard quantities, such as character strings, functions or other kind of non-numerical explanatory variables. The aim of the DBR is to predict the response variable for a new individual $\mathcal{O}_{n+1}$ from the same population, using $(d^2_{n+1,1}, \ldots, d^2_{n+1,n})$, the vector of squared distances from $\mathcal{O}_{n+1}$ to the remaining individuals, as the only available information.

DBR operates as follows. We say that a $n \times r$ matrix $\boldsymbol{X}$, $r \leq n$, is a Euclidean configuration for $\boldsymbol{\Delta}$ if $\boldsymbol{X}$ verifies that the Euclidean distance between its rows i and j is equal to d_{ij}. It is assumed (Euclidean condition) that such a configuration exists for $\boldsymbol{\Delta}$. We denote $\mathbb{R}^r$ by $\mathcal{E}$, the Euclidean space where the rows of $\boldsymbol{X}$ belong. Metric Multidimensional Scaling (see, e.g., [14]) can be used to obtain $\boldsymbol{X}$ from $\boldsymbol{\Delta}$. Then the linear regression of $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ on $\boldsymbol{X}$ is estimated by OLS, giving a r-dimensional estimated regression coefficient $\hat{\boldsymbol{\beta}}$. It can be proven ([13]) that $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is an intrinsic quantity, meaning that it can alternatively be expressed directly as a function of $\boldsymbol{\Delta}$. Therefore $\hat{\boldsymbol{y}}$ is independent of the particular choice of $\boldsymbol{X}$, that is never explicitly computed, and neither is $\hat{\boldsymbol{\beta}}$.

In this paper we present the weighted version of DBR, where each response y_i has a weight $w_i \leq 0$. In principle, this extension requires no more work than replacing OLS by Generalized (Weighted) Least Squares (WLS) in the well-known DBR formulae. Nevertheless, since the derivations require some tricky algebraic details it is worth writing them in full. Subsection 2.1 is a review of Weighted Metric Multidimensional Scaling concepts, with formulae

adapted to our case. Subsection 2.2 contains the derivations and formulae for Weighted DBR.

2.1 Weighted Metric Multidimensional Scaling

In this section we review concepts, notations and formulae of Weighted Metric Scaling, an extension of the usual Metric Multidimensional Scaling, designed to take into account a weight for each individual, either because it has an intrinsic observed multiplicity or due to heteroscedasticity in the response in a DBR (see Section 2.2).

Assume given a set, $\Omega = \{\mathcal{O}_1, \ldots, \mathcal{O}_n\}$, of n individuals where each $\mathcal{O}_i$ is described by an observable $\boldsymbol{z}_i$. Moreover, each individual will have a set of Euclidean coordinates, a vector $\boldsymbol{x}_i$ in some Euclidean space $\mathbb{R}^r$ –precise relationship between both coordinate sets is described below. These vectors are written as $1 \times r$ row vectors, in order to stack them as an $n \times r$ matrix $\boldsymbol{X}$, a Euclidean configuration. In classical applications, Multidimensional Scaling aims at visualizing data, which translates into a requirement that computation of $\boldsymbol{X}$ should be followed by a dimensional reduction (rows of X in $\mathbb{R}^k$, $k < r$), attaining a plane representation ($k = 2$). This constraint needs not be enforced for DBR, where k is usually decided by model stability (bias-variance tradeoff) criteria. We deal with this issue at the end of the next subsection.

The initial step is to compute an $n \times n$ matrix of squared distances:

$$\boldsymbol{\Delta} = \delta_{ij}^2 \equiv \delta^2(\boldsymbol{z}_i, \boldsymbol{z}_j), 1 \leq i, j \leq n,$$

where δ, the distance function, is a semi-metric in Ω. Additionally, each individual $\mathcal{O}_i$ has a positive weight $w_i \in (0, 1)$. The $n \times 1$ weight vector $\boldsymbol{w} = (w_1, \ldots, w_n)'$ is standardized to unit sum, i.e., $\mathbf{1}' \cdot \boldsymbol{w} = 1$, where $\mathbf{1}$ is the $n \times 1$ vector of ones.

One of the key concepts to be considered is that of *projectors*:

$$\boldsymbol{K}_{\boldsymbol{w}} = \mathbf{1} \cdot \boldsymbol{w}'$$

is the $n \times n$ projector, along $\boldsymbol{w}$, on the span $\langle \mathbf{1} \rangle$ of $\mathbf{1}$, and

$$\boldsymbol{J}_{\boldsymbol{w}} = \boldsymbol{I} - \boldsymbol{K}_{\boldsymbol{w}},$$

projects along $\boldsymbol{w}$ on $\langle \mathbf{1} \rangle^{\perp}$. $\boldsymbol{J}_{\boldsymbol{w}}$ can also be named the *centering matrix with respect to* $\boldsymbol{w}$. These projectors are idempotents but not symmetrical nor orthogonal, except in the uniform case, i.e., $\boldsymbol{w} = (1/n)\,\mathbf{1}$, $\boldsymbol{K} = (1/n)\mathbf{1} \cdot \mathbf{1}'$, and $\boldsymbol{J} = \boldsymbol{I} - \boldsymbol{K}$.

The $n \times n$ *inner-products matrix* $\boldsymbol{G}_{\boldsymbol{w}}$ is obtained by

$$\boldsymbol{G}_{\boldsymbol{w}} = -\frac{1}{2}\, \boldsymbol{J}_{\boldsymbol{w}} \cdot \boldsymbol{\Delta} \cdot \boldsymbol{J}_{\boldsymbol{w}}{}'.$$

$\boldsymbol{\Delta}$ can be recovered from $\boldsymbol{G}_{\boldsymbol{w}}$ as

$$\boldsymbol{\Delta} = \mathbf{1} \cdot \boldsymbol{g}_{\boldsymbol{w}} + \boldsymbol{g}'_{\boldsymbol{w}} \cdot \mathbf{1}' - 2\, \boldsymbol{G}_{\boldsymbol{w}}, \tag{1}$$

where $\boldsymbol{g}_{\boldsymbol{w}}$ is a $1 \times n$ row vector containing the (necessarily nonnegative) diagonal entries of $\boldsymbol{G}_{\boldsymbol{w}}$. The *standardized inner-products matrix* is defined as

$$\boldsymbol{F}_{\boldsymbol{w}} = \boldsymbol{D}_{\boldsymbol{w}}^{1/2} \cdot \boldsymbol{G}_{\boldsymbol{w}} \cdot \boldsymbol{D}_{\boldsymbol{w}}^{1/2}, \tag{2}$$

where $\boldsymbol{D}_{\boldsymbol{w}} = \mathrm{diag}(\boldsymbol{w})$ is the diagonal matrix whose diagonal entries are the weights $\boldsymbol{w}$.

Let r be the rank of $\boldsymbol{G}_{\boldsymbol{w}}$. Any $n \times r$ matrix $\boldsymbol{X}_{\boldsymbol{w}}$ such that $\boldsymbol{G}_{\boldsymbol{w}} = \boldsymbol{X}_{\boldsymbol{w}} \cdot \boldsymbol{X}'_{\boldsymbol{w}}$ is a $\boldsymbol{w}$-*centered Euclidean configuration of* $\boldsymbol{\Delta}$, where $\boldsymbol{w}$-centered means that $\boldsymbol{w}' \cdot \boldsymbol{X}_{\boldsymbol{w}} = \mathbf{0}$. It is worth noting that Euclideanarity is an intrinsic or geometric concept, that is, equation (1) is the matrix notation for the set of equalities

$$\delta_{ij}^2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 = \langle \boldsymbol{x}_i, \boldsymbol{x}_i \rangle + \langle \boldsymbol{x}_j, \boldsymbol{x}_j \rangle - 2\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle,$$

which mean that the Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is equal to $\delta(\boldsymbol{z}_i, \boldsymbol{z}_j)$.

Schoenberg's Theorem (see, e.g., [15, Theorem 4]) states that such a decomposition exists if and only if $\boldsymbol{G}_{\boldsymbol{w}}$ is a positive semidefinite (psd) matrix, in which case $\boldsymbol{\Delta}$ is called a *Euclidean distance matrix* (in the Multidimensional Scaling sense). [N.B. $\boldsymbol{G}_{\boldsymbol{w}}$ is psd $\iff$ $\boldsymbol{G} \equiv -\frac{1}{2}\, \boldsymbol{J} \cdot \boldsymbol{\Delta} \cdot \boldsymbol{J}$ is psd, since $\boldsymbol{J}_{\boldsymbol{w}} \cdot \boldsymbol{J} = \boldsymbol{J}_{\boldsymbol{w}}$ and $\boldsymbol{J} \cdot \boldsymbol{J}_{\boldsymbol{w}} = \boldsymbol{J}$].

There are many $\boldsymbol{w}$-centered r-dimensional Euclidean configurations of a given $\boldsymbol{\Delta}$, but any two of them are related by an orthogonal (bijective) transformation. Two Euclidean configurations of $\boldsymbol{\Delta}$ with a different centering, $\boldsymbol{X}_{\boldsymbol{w}}$ and $\boldsymbol{X}$, say, where $\boldsymbol{X}_{\boldsymbol{w}}$ is $\boldsymbol{w}$-centered and $\boldsymbol{X}$ is centered (in the ordinary sense, i.e., $\mathbf{1}' \cdot \boldsymbol{X} = \mathbf{0}$), are related by a translation:

$$\boldsymbol{X}_{\boldsymbol{w}} = \boldsymbol{X} + \mathbf{1} \cdot \boldsymbol{a}, \qquad \text{where } \boldsymbol{a} = -\boldsymbol{w}' \cdot \boldsymbol{X},$$

as is readily checked.

Let $\boldsymbol{S}_{\boldsymbol{w}}$ be the covariances matrix of a $\boldsymbol{w}$-centered Euclidean configuration $\boldsymbol{X}_{\boldsymbol{w}}$,

$$\boldsymbol{S}_{\boldsymbol{w}} = \boldsymbol{X}_{\boldsymbol{w}}{}' \cdot (\boldsymbol{D}_{\boldsymbol{w}} - \boldsymbol{w} \cdot \boldsymbol{w}') \cdot \boldsymbol{X}_{\boldsymbol{w}} = \boldsymbol{X}_{\boldsymbol{w}}{}' \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{X}_{\boldsymbol{w}}.$$

It is easy to prove that the trace of $\boldsymbol{S}_{\boldsymbol{w}}$ is independent of the choice of $\boldsymbol{X}_{\boldsymbol{w}}$ and that

$$\mathrm{tr}(\boldsymbol{S}_{\boldsymbol{w}}) = \mathrm{tr}(\boldsymbol{F}_{\boldsymbol{w}}) = \frac{1}{2} \boldsymbol{w}' \cdot \boldsymbol{\Delta} \cdot \boldsymbol{w}.$$

We name this quantity *Geometric Variability of* $\boldsymbol{\Delta}$ *with respect to* $\boldsymbol{w}$. Observe that $\boldsymbol{S_w}$ is nonsingular because $\boldsymbol{X_w}$ is full rank by columns and $\boldsymbol{D_w}$ is a diagonal matrix.

Assume that a new case $\mathcal{O}_{n+1}$ is available. The only relevant information from that individual is the $1 \times n$ vector $\boldsymbol{d}_{n+1}$ of squared distances from $\mathcal{O}_{n+1}$ to the remaining individuals. The next Proposition allows us to represent this new individual as a r-vector $\boldsymbol{x}_{n+1}$ in the row space of $\boldsymbol{X_w}$, giving the best r-dimensional approximation in the Weighted Least Squares sense to an exact Euclidean configuration of the whole set of $n+1$ individuals with their distances. This result is the generalization for the weighted case of the *Gower's interpolation* or *add-a-point* formula ([16]).

Proposition 1 *The weighted version of the Gower's add-a-point formula is*

$$\hat{\boldsymbol{x}}_{n+1} = \frac{1}{2}\,(\boldsymbol{g_w} - \boldsymbol{d}_{n+1}) \cdot \boldsymbol{D_w} \cdot \boldsymbol{X_w} \cdot \boldsymbol{S_w}^{-1}. \tag{3}$$

Proof. The derivation of (3) proceeds by equating each i-th entry in $\boldsymbol{d}_{n+1}$ to the intended vector quantity in terms of the $1 \times r$ vector $\hat{\boldsymbol{x}}_{n+1}$ and the i-th row of $\boldsymbol{X_w}$:

$$\begin{aligned} \boldsymbol{d}_{n+1,i} &= (\hat{\boldsymbol{x}}_{n+1} - \boldsymbol{x}_i) \cdot (\hat{\boldsymbol{x}}_{n+1} - \boldsymbol{x}_i)' \\ &= \|\hat{\boldsymbol{x}}_{n+1}\|^2 + \|\boldsymbol{x}_i\|^2 - 2\,\hat{\boldsymbol{x}}_{n+1} \cdot \boldsymbol{x}_i', \quad 1 \le i \le n. \end{aligned} \tag{4}$$

Writing as a row vector the results of multiplying each i-th equation by w_i:

$$\boldsymbol{d}_{n+1} \cdot \boldsymbol{D_w} = \|\hat{\boldsymbol{x}}_{n+1}\|^2\, \boldsymbol{w}' + \boldsymbol{g_w} \cdot \boldsymbol{D_w} - 2\,\hat{\boldsymbol{x}}_{n+1} \cdot \boldsymbol{X_w}' \cdot \boldsymbol{D_w}. \tag{5}$$

Multiplying (5) on the right by $\boldsymbol{X_w}$ and collecting terms:

$$(\boldsymbol{g_w} - \boldsymbol{d}_{n+1}) \cdot \boldsymbol{D_w} \cdot \boldsymbol{X_w} = 2\,\hat{\boldsymbol{x}}_{n+1} \cdot \boldsymbol{S_w}.$$

Finally, since $\boldsymbol{S_w}$ is nonsingular, we obtain (3). □

2.2 Weighted Distance-Based Regression

Given a $n \times 1$ weight vector $\boldsymbol{w}$ and an $n \times 1$ $\boldsymbol{w}$-centered numerical vector $\boldsymbol{y}$ (that is, $\boldsymbol{w}'\boldsymbol{y} = 0$), the Weighted DBR of response $\boldsymbol{y}$ with weights $\boldsymbol{w}$ and predictor matrix $\boldsymbol{\Delta}$, an $n \times n$ square distances matrix, is defined as the WLS regression of $\boldsymbol{y}$, on a $\boldsymbol{w}$-centered Euclidean configuration of $\boldsymbol{\Delta}$, $\boldsymbol{X_w}$, with weights $\boldsymbol{w}$.

A key result, Proposition 2, states that the hat matrix for this regression,

$$\boldsymbol{H}_{\boldsymbol{w}} = \boldsymbol{X}_{\boldsymbol{w}} \cdot (\boldsymbol{X}_{\boldsymbol{w}}{}' \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{X}_{\boldsymbol{w}})^{-1} \cdot \boldsymbol{X}_{\boldsymbol{w}}{}'. \tag{6}$$

is an intrinsic quantity, meaning that it can be expressed directly as a function of the distances or, equivalently, the inner products. To prove this result we will need:

Lemma 1 *The Moore-Penrose pseudo-inverse of the standardized inner-products matrix, $\boldsymbol{F}_{\boldsymbol{w}}$, defined in (2), can be expressed in terms of a $\boldsymbol{w}$-centered Euclidean configuration $\boldsymbol{X}_{\boldsymbol{w}}$ as:*

$$\boldsymbol{F}_{\boldsymbol{w}}{}^{+} = \boldsymbol{D}_{\boldsymbol{w}}^{1/2} \cdot \boldsymbol{X}_{\boldsymbol{w}} \cdot (\boldsymbol{X}_{\boldsymbol{w}}{}' \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{X}_{\boldsymbol{w}})^{-2} \cdot \boldsymbol{X}_{\boldsymbol{w}}{}' \cdot \boldsymbol{D}_{\boldsymbol{w}}^{1/2}. \tag{7}$$

Proof. A direct computation, checking the properties characterizing the Moore-Penrose pseudo-inverse of a given matrix $\boldsymbol{M}$: To this end, a matrix $\boldsymbol{N}$ has to satisfy: 1) $\boldsymbol{M} \cdot \boldsymbol{N} \cdot \boldsymbol{M} = \boldsymbol{M}$, 2) $\boldsymbol{N} \cdot \boldsymbol{M} \cdot \boldsymbol{N} = \boldsymbol{N}$, 3) both $\boldsymbol{M} \cdot \boldsymbol{N}$ and $\boldsymbol{N} \cdot \boldsymbol{M}$ are symmetric. □

Straightforward algebra from Lemma 1, shows:

Proposition 2 *The hat matrix (6) is equal to:*

$$\boldsymbol{H}_{\boldsymbol{w}} = \boldsymbol{G}_{\boldsymbol{w}} \cdot \left(\boldsymbol{D}_{\boldsymbol{w}}{}^{1/2} \cdot \boldsymbol{F}_{\boldsymbol{w}}{}^{+} \cdot \boldsymbol{D}_{\boldsymbol{w}}{}^{1/2}\right)$$

and $\hat{\boldsymbol{y}} = \boldsymbol{H}_{\boldsymbol{w}} \cdot \boldsymbol{y}$ is

$$\hat{\boldsymbol{y}} = \boldsymbol{G}_{\boldsymbol{w}} \cdot \left(\boldsymbol{D}_{\boldsymbol{w}}{}^{1/2} \cdot \boldsymbol{F}_{\boldsymbol{w}}{}^{+} \cdot \boldsymbol{D}_{\boldsymbol{w}}{}^{1/2}\right) \cdot \boldsymbol{y}. \tag{8}$$

The next result allows to evaluate the predicted Y for a new case $\mathcal{O}_{n+1}$, given its $\boldsymbol{d}_{n+1}$ vector, defined above, in the paragraph before Proposition 1.

Proposition 3

$$\hat{y}_{n+1} = \frac{1}{2} \left(\boldsymbol{g}_{\boldsymbol{w}} - \boldsymbol{d}_{n+1}\right) \cdot \left(\boldsymbol{D}_{\boldsymbol{w}}{}^{1/2} \cdot \boldsymbol{F}_{\boldsymbol{w}}{}^{+} \cdot \boldsymbol{D}_{\boldsymbol{w}}{}^{1/2}\right) \cdot \boldsymbol{y}. \tag{9}$$

Proof. Normal equations for the WLS regression of $\boldsymbol{y}$ on $\boldsymbol{X}_{\boldsymbol{w}}$ yield the estimated $\hat{\boldsymbol{\beta}}$ as:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_{\boldsymbol{w}}{}' \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{X}_{\boldsymbol{w}})^{-1} \cdot \boldsymbol{X}_{\boldsymbol{w}}{}' \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{y} = \boldsymbol{S}_{\boldsymbol{w}}{}^{-1} \cdot \boldsymbol{X}_{\boldsymbol{w}}{}' \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{y}.$$

The prediction $\hat{y}_{n+1}$ for $\mathcal{O}_{n+1}$ is

$$\hat{y}_{n+1} = \hat{\boldsymbol{x}}_{n+1} \cdot \hat{\boldsymbol{\beta}},$$

where $\hat{\boldsymbol{x}}_{n+1}$ is the result of Gower's interpolation (3). That is,

$$\hat{y}_{n+1} = \frac{1}{2}\left(\boldsymbol{g}_{\boldsymbol{w}} - \boldsymbol{d}_{n+1}\right) \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{X}_{\boldsymbol{w}} \cdot {\boldsymbol{S}_{\boldsymbol{w}}}^{-2} \cdot {\boldsymbol{X}_{\boldsymbol{w}}}' \cdot \boldsymbol{D}_{\boldsymbol{w}} \cdot \boldsymbol{y}.$$

Comparison with (7) yields (9). □

Equations (8) and (9) are the core of Weighted DBR. Observe that it is a linear regression in the space $\mathcal{E}$ where the Euclidean configuration $\boldsymbol{X}_{\boldsymbol{w}}$ is included. In practice this configuration is not explicitly calculated.

It is also remarkable that Weighted DBR reproduces the results of WLS: if we start from a $n \times r$ matrix $\boldsymbol{X}$ of r continuous independent variables corresponding to n individuals (with weights given by $\boldsymbol{w}$) and we define $\boldsymbol{\Delta} = (d_{i,j}^2)$, d_{ij} being the Euclidean distance between rows i and j of $\boldsymbol{X}$, then $\hat{\boldsymbol{y}}_{WDBR} = \hat{\boldsymbol{y}}_{WLS}$ and $\hat{y}_{n+1,WDBR} = \hat{y}_{n+1,WLS}$, because $\boldsymbol{X}$ is trivially a Euclidean configuration for $\boldsymbol{\Delta}$. A particular example is when the i-th row of $\boldsymbol{X}$ is (x_i, x_i^2, x_i^3), $x_i \in \mathbb{R}$. Then doing the cubic weighted regression of y_i over x_i is equivalent to fitting Weighted DBR with distances $d(x_i, x_j) = \|(x_i, x_i^2, x_i^3) - (x_j, x_j^2, x_j^3)\|_2$.

There is a technicality (related with dimensional reduction) worth discussing: the rank r of the hat-matrix in (8), as in an ordinary linear regression, is equivalent to the number of linearly independent linear predictors. Since for n cases, depending on the metric chosen, r can be as high as $n-1$, giving an overdetermined model with unstable predictions, a sensible procedure is to replace the pseudo-inverse $\boldsymbol{F}_{\boldsymbol{w}}^{+}$ with a lower-rank approximation. This can be easily implemented by the Singular Value Decomposition which, by the Schmidt-Eckart-Young Theorem ([17]), gives the best ℓ^2 approximation of any given rank k, $1 \leq k \leq r$. A cross-validation statistic can then be used to select a suitable k. We use a *leave-one-out* scheme to compute a cross-validation Mean Square Prediction Error (MSPE):

$$\text{MSPE} = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i^{-i})^2,$$

where $\hat{y}_i^{-i}$ is the i-th prediction, as derived from the $n-1$ remaining cases.

2.3 Weighted DBR: An example from Insurance

As an illustration, we apply Weighted DBR to a well-known dataset of damage claim amounts in a car insurance portfolio (Table 1 in [18]). These data have been studied by many authors using diverse methods (see, e.g. [19], [20], [21]).

This dataset is a cross-tabulation according to risk profiles, which are to be

used as predictors. For each i-th nonempty cell ($1 \leq i \leq n = 123$) the response variable y_i is the average claim amount of the N_i individual claims in the cell.

The set $\boldsymbol{z}$ of predictors consists of three risk factors:

- *Policyholder's Age* (expressed in years) is a continuous numerical measurement which, already in the original data, has been discretized into 8 classes: 17–20, 21–24, 25–29, 30–34, 35–39, 40–49, 50–59, 60+. In order to process quantitatively this variable, we use the following class marks: 18.5, 22.5, 27.0, 32.0, 37.0, 44.5, 54.5, 65.0.
- *Car Group* is a categorical variable, with 4 levels, labelled A, B, C, D.
- *Vehicle Age* (expressed in years) is a continuous numerical measurement, discretized into 4 classes: 0–3, 4–7, 8–9, 10+. In order to process quantitatively this variable, we use the following class marks: 1.5, 5.5, 8.5, 12.0.

As usual, for these data, the natural weights for all prediction models, including Weighted DBR are the cell frequencies N_i, which have been used to compute the average claim amounts y_i.

The first step in the treatment of these data by Weighted DBR is the choice of a suitable metric. In principle it is possible to tailor a metric to reflect specific information on predictors and on how their proximity relates to the particular prediction under study. Here it is sufficient to utilize an omnibus metric function which satisfies the Euclidean condition. One very popular such metric for mixtures of numerical continuous, categorical and binary predictor variables is the one based on Gower's general similarity coefficient (see [22]), which for two p-dimensional vectors $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ is equal to

$$s_{ij} = \frac{\sum_{h=1}^{p_1} \left(1 - \left|z_{ih} - z_{jh}\right| / R_h\right) + a + \alpha}{p_1 + (p_2 - d) + p_3}, \tag{10}$$

where $p = p_1 + p_2 + p_3$, p_1 is the number of continuous variables, a and d are the number of positive and negative matches, respectively, for the p_2 binary variables, and α is the number of matches for the p_3 multi-state categorical variables. R_h is the range of the h-th continuous variable. The squared distance is computed as:

$$d^2(\boldsymbol{z}_i, \boldsymbol{z}_j) = 1 - s_{ij}. \tag{11}$$

Gower [22] proves that (11) satisfies the Euclidean condition. In our case, $p_1 = 2$, $p_2 = 0$, $p_3 = 1$.

The next step is to apply the Weighted DBR formulae, taking into account the technicality on the rank choice explained at the end of Subsection 2.2. The results of Weighted DBR for our data, with Gower's metric, are: the maximum rank is $r = 13$, and the optimal cross-validation MSPE value is 1431.8, for $k = 9$. Left panel in Figure 1 shows the MSPE as a function of the rank.

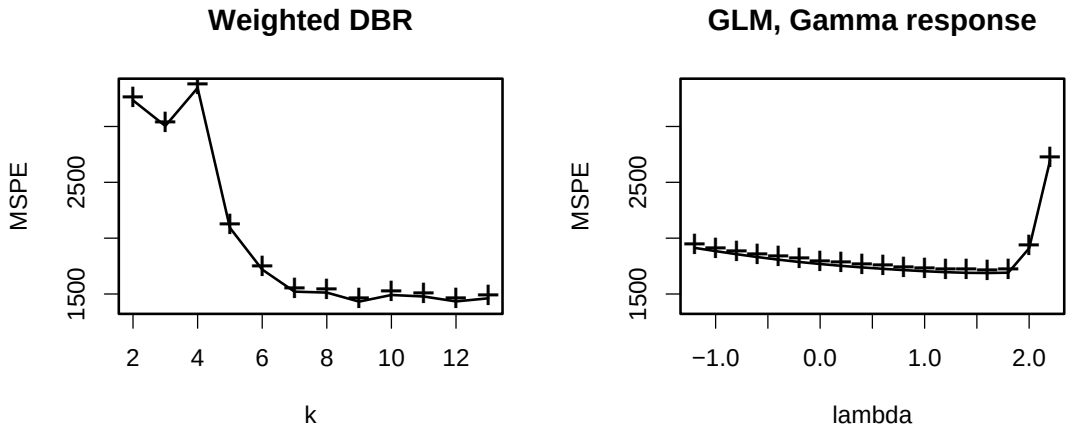

Fig. 1. Cross-validation MSPE statistics for Weighted DBR (left) as a function of rank k, and for GLM with Gamma response and Box-Cox link family, for several λ values.

In order to compare the Weighted DBR results with those of a more conventional Generalized Linear Model (GLM) treatment, we consider a Gamma response distribution and a Box-Cox family of link functions, following [19, p. 377], [21, p. 204–209], but for the sake of a closer parallelism, we have recomputed the GLM's treating the predictors *Policyholder's Age* and *Vehicle Age* as described above, keeping their quantitative character, instead of as qualitative variables coded with dummy indicators as these authors do. Right panel in Figure 1 shows the cross-validation MSPE statistic as a function of the exponent λ in the Box-Cox link function ($\lambda = 0$ corresponds to the logarithmic link). The optimal cross-validation MSPE value is 1687.9, for $\lambda = 1.6$. As a final comparison, considering a Gaussian response with identity link, we obtain MSPE = 1717.0. We observe that for this data set Weighted DBR gives the best result among the considered alternatives.

3 Local Linear Distance-Based Regression

Let $(\chi_i; y_i),\ i = 1, \ldots, n$, be a random sample of $(\boldsymbol{\chi}, Y)$, $Y \in \mathbb{R}$, $\boldsymbol{\chi} : [a, b] \to \mathbb{R}$. We want to estimate $m(\chi) = E(Y|\boldsymbol{\chi} = \chi)$ by a local linear regression around χ and we are doing that using Weighted DBR. We consider the weights

$$w_i(\chi) = K(d_1(\chi, \chi_i)/h) / \sum_{j=1}^{n} K(d_1(\chi, \chi_j)/h)$$

where d_1 is a semi-metric between functions. Let $\boldsymbol{\Delta}_2$ be the matrix of squared distances between functions defined from a possible different semi-metric d_2.

We fit a Weighted DBR starting from the initial elements

$$\boldsymbol{\Delta}_2 = (d_2(\chi_i, \chi_j)^2)_{i=1..n, j=1..n},\ \boldsymbol{y} = (y_i)_{i=1..n},\ \text{ and } \boldsymbol{w} = (w_i(\chi))_{i=1..n}.$$

We consider a new individual $\mathcal{O}_{n+1}$ where the functional predictor is χ and we compute its squared distances to the other individuals χ_i:

$$\boldsymbol{d}_{2,n+1} = (d_2(\chi, \chi_1)^2, \cdots, d_2(\chi, \chi_n)^2).$$

Then we use equation (9) to obtain the *Local Linear DBR* estimator of $m(\chi)$:

$$\hat{m}_{LLDBR}(\chi) = \hat{y}_{n+1}.$$

Let us remark some important points. There are two semi-metrics involved in the local linear distance-based estimation: one of them, d_1, is used to compute the weight of observation χ_i around the function χ where the regression function is estimated, and the other, d_2, defines the distances between observations for computing the DBR. The semi-metrics d_1 and d_2 can coincide or not. Observe that the local linear distance-based estimator of $m(\chi)$ is really a local linear estimator in the space $\mathcal{E}$ where the semi-metric d_2 is a Euclidean distance.

Assume that d_1 and d_2 coincide and that they are the Euclidean distance in $L_2([a,b])$, that is, $d_1 = d_2 = d_0^{deriv}$. Then the local linear distance-based estimator $\hat{m}_{LLDBR}(\chi)$ coincide with the linear local estimator $\hat{m}_{LL}(\chi)$ proposed by Baíllo and Grané in [7]. Assume now that $d_2(\chi, \gamma) = 0$ for all functions χ and γ. Then the local linear distance-based estimator $\hat{m}_{LLDBR}(\chi)$ fits locally a constant around χ and then it coincides with the kernel estimator $\hat{m}_K(\chi)$ introduced by Ferraty and Vieu in [2].

Let K be the uniform kernel and assume that $h > \max_{i,j}(d_1(\chi_i, \chi_j))$. Then a (global) DBR is fitted, that is a linear regression fit in the space $\mathcal{E}$ where the semi-metric d_2 is a Euclidean distance.

The local linear distance-based estimation is also valid for predictors that are no functional data. For instance, it is valid for multivariate continuous data ($x_i \in \mathbb{R}^p$), mixed data (multivariate x_i with some components being continuous and other being qualitative), textual data or any other kind of data for which we are able to compute distances between individuals. Consider, for instance, that $x_i \in \mathbb{R}$, $d_1(x_i, x_j) = |x_i - x_j|$, $d_2(x_i, x_j) = \|(x_i, x_i^2, x_i^3) - (x_j, x_j^2, x_j^3)\|$. Then the estimator $\hat{m}_{LLDBR}(x)$ coincides with fitting a local cubic polynomial regression (see the end of Subsection 2.2).

Observe that $\hat{m}_{LLDBR}$ is a *linear smoother* (as defined in [23]) in the sense that $\hat{\boldsymbol{y}} = \boldsymbol{S} \cdot \boldsymbol{y}$, where $\hat{\boldsymbol{y}} = (\hat{y}_i)_{i=1..n}$ and $\boldsymbol{S}_{n\times n}$, the *smoothing matrix*, only depends on distances d_1 and d_2 between observed functions χ_i. This property allows

the definition of the *effective degrees of freedom* of $\hat{m}_{LLDBR}$ (as trace($\boldsymbol{S}$); see [24] or [25]), a quantity that could be useful in practice to quantify the degree of smoothing of $\hat{m}_{LLDBR}$, and the *effective kernel* for estimating $m(\chi_i)$, the i-th row of $\boldsymbol{S}$. In particular, the expression of the i-th row of $\boldsymbol{S}$ can be derived from the equation (8). Let $w_{ij} = K(d_1(\chi_i, \chi_j)/h)/\sum_{l=1}^{n} K(d_1(\chi_i, \chi_l)/h)$, let $\boldsymbol{w}_i = (w_{ij})_{j=1..n}$ and let $\boldsymbol{\Delta}_2$ be as defined at the beginning of this section. Let $\boldsymbol{D}_{\boldsymbol{w}_i} = \mathrm{diag}(\boldsymbol{w}_i)$, let $\boldsymbol{J}_{\boldsymbol{w}_i} = \boldsymbol{I} - \boldsymbol{1} \cdot \boldsymbol{w}_i'$, let $\boldsymbol{G}_{\boldsymbol{w}_i} = -\frac{1}{2}\, \boldsymbol{J}_{\boldsymbol{w}_i} \cdot \boldsymbol{\Delta}_2 \cdot \boldsymbol{J}_{\boldsymbol{w}_i}{}'$, let $\boldsymbol{g}_{\boldsymbol{w}_i}$ be the i-th row of $\boldsymbol{G}_{\boldsymbol{w}_i}$, let $\boldsymbol{F}_{\boldsymbol{w}_i} = \boldsymbol{D}_{\boldsymbol{w}_i}^{1/2} \cdot \boldsymbol{G}_{\boldsymbol{w}_i} \cdot \boldsymbol{D}_{\boldsymbol{w}_i}^{1/2}$ and let $\boldsymbol{F}_{\boldsymbol{w}_i}{}^{+}$ be the Moore-Penrose pseudo-inverse of $\boldsymbol{F}_{\boldsymbol{w}_i}$. Then, the i-th row of $\boldsymbol{S}$ is

$$\boldsymbol{g}_{\boldsymbol{w}_i} \cdot \left(\boldsymbol{D}_{\boldsymbol{w}_i}{}^{1/2} \cdot \boldsymbol{F}_{\boldsymbol{w}_i}{}^{+} \cdot \boldsymbol{D}_{\boldsymbol{w}_i}{}^{1/2} \right).$$

3.1 A real data example: Spectrometric Data

We consider the *Spectrometric Data* [2] already introduced in Section 1. Remember that there are 215 samples of chopped meat and that for each case, the spectrometric function χ and the sample fat content y are available. The goal is to predict the fat content y from χ.

Following Section 7.2 in [2] we divide the sample in a training sample (the first 160 cases) and a test sample (the last 55 cases). The performance of different functional prediction methods is measured by the empirical mean square prediction error in the sample test: $MSPE = (1/55)\sum_{i=161}^{215}(\hat{y}_i - y_i)^2$.

Ferraty and Vieu [2] use three functional predictors for this data set: non-parametric estimators of conditional expectation (functional kernel estimator as $\hat{m}_K(\chi)$), conditional mode and conditional median. The implementation of these estimators allows a variable bandwidth h based on k-nearest neighbours, where k is locally selected by cross-validation. The authors recommend to use the semi-metric based on the second order derivatives (d_2^{deriv}). We have used the R routines accompanying [2] (the script `npfda-specpredRS.txt` to be specific) to recreate the results included in the book. The numbers we have obtained are shown in Table 1 with the label FV2006.

In order to have results that we can directly compare with our proposals, we have computed the functional kernel estimators with fixed bandwidth selected by cross-validation and based on the semi-metrics d_r^{deriv}, $r = 1, 2, 3$. We have used the R function from [2] `funopare.kernel.cv`. The results are included in Table 1 with the label Kernel.FV.

We have implemented the local linear DBR with automatic selection of the bandwidth by cross-validation. The usual way of implementing cross-validation has been modified as follows. Usually it is not possible to check the performance of a candidate bandwidth h being lower than $\max_i \min_j d_1(\chi_i, \chi_j) =$

$\min_j d_1(\chi_{i^*}, \chi_j)$ because in this case there are not enough data in the ball centered at χ_{i^*} with radius h to fit the DBR. So for an observation χ_i having less than 3 neighbours at distance h, we enlarge h to h_i allowing to include 3 neighbours in the ball centered at χ_{i^*} with radius h_i. So our implementation is with partially variable bandwidth.

When doing Weighted DBR internal steps, we use full rank $\boldsymbol{F}^+_{\boldsymbol{w}}$ matrices in equations (8) and (9) instead of choosing this rank by cross-validation (see comments on that at the end of Subsection 2.2). We consider that choosing h minimizing the MSPE is an indirect way to control the rank of $\boldsymbol{F}^+_{\boldsymbol{w}}$ because the value of h determines the number of observations involved in each Weighted DBR estimation.

An alternative implementation of functional kernel estimators is possible using local linear DBR by selection $d_1 = d_r^{deriv}$, $r = 1, 2, 3$, and $d_2 \equiv 0$. The results are included in Table 1 with the label Kernel.LLDBR. The results do not coincide with those obtained using the function `funopare.kernel.cv` because the different way of bandwidth selection.

Finally we also show in Table 1 the results obtained by local linear DBR for different combinations of distances d_1 and d_2, all of them using semi-metrics based on derivatives. First we fix d_2 equal to the Euclidean distance between the original functions ($d_2 = d_0^{deriv}$) and use $d_1 = d_r^{deriv}$, $r = 1, 2, 3$. This way we do local linear regression in the space of the original functions for different semi-metrics defining neighborhoods in this space. The case $d_1 = d_0^{deriv}$ and $d_2 = d_0^{deriv}$ corresponds to the local linear estimator proposed in [7]. The case $d_1 = d_2^{deriv}$ and $d_2 = d_0^{deriv}$ represents an improvement on the kernel method (compare with the row labeled Kernel.LLDBR d_2^{deriv}) because now a local linear regression is fitted instead computing a local average. The best fitting is obtained when using $d_1 = d_2^{deriv}$ and $d_2 = d_2^{deriv}$: local linear regression in the space of second derivatives. This choice of distances d_1 and d_2 is also the most natural one taken into account the recommendations of Section 7.2 in [2].

4 Conclusions

We have presented the local linear DBR estimator of $m(\chi)$, a nonparametric method based on Weighted DBR. This method is very flexible, including as a particular case the local polynomial regression for real predictor variables. Moreover it gives good results in practice. So we consider that this proposal is a satisfactory answer to *Open question 5* in [2].

There are practical and theoretical matters that deserve further attention.

Table 1
Mean square prediction error (MSPE) for different functional predictors.

Functional predictor	MSPE	Functional predictor	MSPE
FV2006 Cond. Expect.	1.92	Kernel.LLDBR d_0^{deriv}	52.08
FV2006 Cond. Mode.	2.94	Kernel.LLDBR d_1^{deriv}	6.85
FV2006 Cond. Median.	4.84	Kernel.LLDBR d_2^{deriv}	3.52
Kernel.FV d_0^{deriv}	139.36	$d_1 = d_0^{deriv}$, $d_2 = d_0^{deriv}$	7.94
Kernel.FV d_1^{deriv}	11.93	$d_1 = d_1^{deriv}$, $d_2 = d_0^{deriv}$	2.12
Kernel.FV d_2^{deriv}	5.37	$d_1 = d_2^{deriv}$, $d_2 = d_0^{deriv}$	1.43
		$d_1 = d_1^{deriv}$, $d_2 = d_1^{deriv}$	2.91
		$d_1 = d_2^{deriv}$, $d_2 = d_2^{deriv}$	1.03

For instance, it is known that for large samples distance-based methods have a high computational cost. For a moderately large sample size ($n \sim 10^4$) out-of-core algorithms analogous to those presented in [26] for DB-PLS regression provide a workable path of solution. Larger sizes require special subsampling. On the other way, the asymptotic properties of the local linear DBR should be studied. Two questions related with the smoothing parameter choice need to be investigated: we could take advantage of the proposed estimator $\hat{m}_{DBLL}$ being a linear smoother (see Section 7.10 in [24] or Section 5.3 in [25], for instance); a variable bandwidth estimator could be defined based on k-Nearest Neighbours following the ideas of Chapter 7 in [2].

In addition to its role underlying Local Linear Functional DBR presented in this paper, Weighted DBR can be directly applied to linear regression with heteroscedastic responses and, cast within an iterative weighted least squares scheme, provides a basis for DB versions of Generalized Linear Models.

Acknowledgments

Work supported in part by the Spanish Ministerio de Educación y Ciencia and FEDER grant MTM2006-09920.

References

[1] J.O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, second edition, 2005.

[2] F. Ferraty and P. Vieu. *Non parametric functional data analysis. Theory and practice*. Springer, 2006.

[3] M. Davidian, X. Lin, and J.L. Wang. Introduction: Emerging issues in longitudinal and functional data analysis. *Statistica Sinica*, 14:613–614, July 2004.

[4] W. González-Manteiga and P. Vieu. Editorial: Statistics for functional data. *Comput. Stat. Data Anal.*, 51:4788–4792, 2007.

[5] M. J. Valderrama. Editorial: An overview to modelling functional data. *Computational Statistics*, 22:331–334, September 2007. Special issue on Modelling Functional Data in Practice.

[6] F. Ferraty, A. Mas, and P. Vieu. Nonparametric regression on functional data: Inference and practical aspects. *Australian and New Zeland J. Stats.*, 49:267–286, 2007.

[7] A. Baíllo and A. Grané. Local linear regression for functional predictor and scalar response. *Journal of Multivariate Analysis*, 2008. (Accepted for publication).

[8] A. Berlinet, A. Elamine, and A. Mas. Local linear regression for functional data. Technical report, ArXiv e-prints, October 2007.

[9] J. Barrientos-Marin. *Some Practical Problems of Recent Nonparametric Procedures: Testing, Estimation, and Application.* PhD thesis, Univ. de Alicante, 2007.

[10] C. M. Cuadras. Distance analysis in discrimination and classification using both continuous and categorical variables. In Y. Dodge, editor, *Statistical Data Analysis and Inference*, pages 459–473, Amsterdam, The Netherlands, 1989. North-Holland Publishing Co.

[11] C.M. Cuadras and C. Arenas. A distance-based regression model for prediction with mixed data. *Communications in Statistics A. Theory and Methods*, 19:2261–2279, 1990.

[12] C. M. Cuadras, C. Arenas, and J. Fortiana. Some computational aspects of a distance-based model for prediction. *Communications in Statistics B. Simulation and Computation*, 25:593–609, 1996.

[13] E. Boj, M. M. Claramunt, and J. Fortiana. Selection of predictors in distance-based regression. *Communications in Statistics A. Theory and Methods*, 36:87–98, 2007.

[14] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications (2nd ed)*. Springer-Verlag, New York, 2005.

[15] J. C. Gower and P. Legendre. Metric and euclidean properties of dissimilarity coefficients. *Journal of Classification*, 3:5–48, 1986.

[16] J. C. Gower. Adding a point to vector diagrams in multivariate analysis. *Biometrika*, 55:582–585, 1968.

[17] G. W. Stewart. On the early history of the singular values decomposition. *SIAM Review*, 35:551–566, 1993.

[18] L. A. Baxter, S. M. Coutts, and G. A. F. Ross. Applications of linear models in motor insurance. In *Transactions of the 21st International Congress of Actuaries*, pages 11–29, 1980. The actual data may be retrieved from the `OzDASL` library at `www.StatSci.org`, under the name "British Car Insurance Claims for 1975"or within the `forward` package in `R`, as dataset `carinsuk`.

[19] P. McCullagh and J. A. Nelder. *Generalized Linear Models (2nd ed)*. Chapman and Hall, London, 1989.

[20] A. C. Cameron and F. A. G. Windmeijer. An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77:329–342, 1997.

[21] A. Atkinson and M. Riani. *Robust Diagnostic Regression Analysis*. Springer-Verlag, New-York, 2000.

[22] J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27:857–874, 1971.

[23] A. Buja, T. Hastie, and R. Tibshirani. Linear smoothers and additive models. *The Annals of Statistics*, 17:453–555, 1989.

[24] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer, 2001.

[25] L. Wasserman. *All of Nonparametric Statistics*. Springer, New York, 2006.

[26] E. Boj, A. Grane, J. Fortiana, and M. M. Claramunt. Implementing pls for distance-based regression: computational issues. *Computational Statistics*, 22:237–248, 2007.