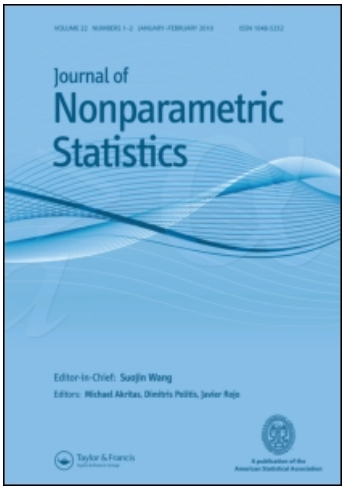

Journal of Nonparametric Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713645758

Locally modelled regression and functional data

J. Barrientos-Marin[a]; F. Ferraty[b]; P. Vieu[b]
[a] University of Alicante, Alicante, Spain [b] Institut de Mathématiques de Toulouse, Equipe Statist. Prob., Université P. Sabatier, Toulouse, France

First published on: 18 November 2009

To cite this Article Barrientos-Marin, J. , Ferraty, F. and Vieu, P.(2010) 'Locally modelled regression and functional data', Journal of Nonparametric Statistics, 22: 5, 617 — 632, First published on: 18 November 2009 (iFirst)
To link to this Article: DOI: 10.1080/10485250903089930
URL: http://dx.doi.org/10.1080/10485250903089930

Journal of Nonparametric Statistics
Vol. 22, No. 5, July 2010, 617–632

Taylor & Francis
Taylor & Francis Group

Locally modelled regression and functional data

J. Barrientos-Marin[a], F. Ferraty[b]* and P. Vieu[b]

[a]*University of Alicante, Alicante, Spain;* [b]*Institut de Mathématiques de Toulouse, Equipe Statist. Prob., Université P. Sabatier, 31062 Toulouse, France*

(*Received 18 November 2008; final version received 29 May 2009*)

The general framework of this paper deals with the nonparametric regression of a scalar response on a functional variable (i.e. one observation can be a curve, surface, or any other object lying into an infinite-dimensional space). This paper proposes to model local behaviour of the regression operator (i.e. the link between a scalar response and an explanatory functional variable). To this end, one introduces a functional approach in the same spirit as local linear ideas in nonparametric regression. The main advantage of this functional local method is to propose an explicit expression of a kernel-type estimator which makes its computation easy and fast while keeping good predictive performance. Asymptotic properties are stated, and a functional data set illustrates the good behaviour of this functional locally modelled regression method.

Keywords: functional data; locally modelled regression; functional nonparametric statistics; rates of convergence; spectrometric curves

AMS Subject Classification: 62G05

Nonparametric statistical models have taken an important place in statistical science. At the same time there are an increasing number of situations arising from different fields of applied sciences in which the data are of a functional nature (i.e. one observation can be a curve, surface, etc.). A lot of functional data sets have been studied in the recent literature, dealing with chemometrics (Frank and Friedman 1993; Ferraty and Vieu 2002; Abraham, Cornillon, Matzner-Löber, and Molinari 2003), radar waveforms (Hall, Poskitt, and Presnell 2001; Dabo-Niang, Ferraty, and Vieu 2004), biometrics (Ramsay, Altman, and Bock 1994; Gasser, Hall, and Presnell 1998), and physiology (Abramovitch, Antoniadis, Sapatinas, and Vidakovic 2004; Abramovitch and Angelini 2006; Antoniadis and Sapatinas 2007). Many other examples can be found in the books (Ramsay and Silverman 2002, 2005; Ferraty and Vieu 2006; Dabo-Niang and Ferraty 2008) and in the special issues devoted to this topic by several statistical journals (Davidian, Lin, and Wang 2004; González Manteiga and Vieu 2007; Valderrama 2007). Such data are called functional data in the sense that they come from observations of a functional variable. The combination of the nonparametric models with the functional data leads us to the problem of the functional nonparametric statistics, which is a very recent field of investigation (Ferraty and Vieu 2006).

*Corresponding author. Email: ferraty@cict.fr

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250903089930
http://www.informaworld.com

This work deals with the functional regression setting when one wishes to predict a response Y from an explanatory functional variable $\mathcal{X}$. In addition, only regularity constraints will be assumed with respect to the regression operator, leading naturally to nonparametric modelling. So, the general problem of this work is the functional nonparametric regression. On the one hand, some works dealing with nonparametric functional regression exist already in the recent literature (see, for instance, Ferraty and Vieu (2002, 2006) for more recent and deeper developments). This functional nonparametric regression method is essentially based on an extension of the well-known Nadaraya–Watson kernel estimator of the regression (Nadaraya 1964; Watson 1964) to the case of explanatory functional variables (see also the recent work of Burba, Ferraty and Vieu (2009) which investigates the k-nearest-neighbours (kNN) estimator). On the other hand, local linear ideas have been developed in the regression context for univariate or multivariate explanatory variables (see Wand and Jones (1995) for an overview on this topic). Very recently, local linear regression models have been investigated when the explanatory variable is a functional predictor (see Aneiros-Pérez, Cao, and Vilar-Fernández (2008); Boj, Delicado, and Fortiana (2008); and Baíllo and Grané (2009) for more details). This is a direct extension of the functional nonparametric regression model (this latter can be viewed as a functional local constant method).

The aim of this work consists of proposing a new local modelling approach when one regresses a scalar response on an explanatory functional variable. This local method is a trade-off between the functional nonparametric regression and the functional local linear one. This new functional local approach leads us to quite a simple kernel estimator; its fast computation and its good predictive performance make this method very attractive (especially when one has to deal with a large data set). Moreover, the simplicity of this locally modelled estimator allows us to get the almost complete rate of convergence (whereas Baíllo and Grané (2009) states the rate of convergence of the conditional mean squared error, in probability). Section 1 describes the position of the problem. Section 2 introduces our functional locally modelled regression method. Its behaviour in practice is illustrated in Section 3 by means of a standard spectrometric data set. Moreover, the combination of this local approach with the functional nonparametric regression allows us to get similar predictive performance to the one obtained with the functional local linear method (with a much lower computational cost). Some asymptotic properties are given in Section 4. Finally, the reader interested by the theoretical developments will find detailed proofs in the appendices.

1. Position of the problem

This paper focuses on the nonparametric estimation of the regression operator defined by

$$Y = m(\mathcal{X}) + \varepsilon, \quad \text{with } \mathbb{E}(\varepsilon|\mathcal{X}) = 0,$$

where the explanatory variable $\mathcal{X}$ is valued in some infinite-dimensional space $\mathcal{H}$ and Y is a scalar response. To do that, one way (Ferraty and Vieu 2006) includes the use of a functional kernel estimator which is an extension to this functional framework of the Nadaraya–Watson kernel estimator. Based on n pairs $(\mathcal{X}_i, Y_i)_{i=1,\dots,n}$ identically and independently distributed as $(\mathcal{X}, Y)$, the functional kernel estimator is defined as follows:

$$m_{\mathrm{LCRE}}(\chi) = \frac{\sum_{i=1}^{n} Y_i K(h^{-1}|\delta(\chi, \mathcal{X}_i)|)}{\sum_{i=1}^{n} K(h^{-1}|\delta(\chi, \mathcal{X}_i)|)},$$

where K is a standard univariate asymmetrical kernel function, $\delta(\cdots)$ locates one element of $\mathcal{H}$ with respect to another one, and h is the so-called bandwidth which plays the role of a smoothing parameter. This kernel estimator $\widehat{m}(\chi)$ can be seen as the solution of the minimisation

problem ($\mathcal{P}_1$):

$$\min_a C_\chi(a) \text{ with } C_\chi(a) = \sum_{i=1}^{n} (Y_i - a)^2 K(h^{-1}|\delta(\chi, \mathcal{X}_i)|),$$

since it is easy to check that the derivative of C_χ vanishes at $a = m_{\text{LCRE}}(\chi)$. Actually, $m_{\text{LCRE}}(\chi)$ is a local weighted average of the Y_i's and can be seen as a *local constant* regression estimator (LCRE) because it approximates locally the Y_i's with a constant. One way to increase the flexibility of the functional nonparametric regression estimator is to use a local approximation which is more accurate than a constant one. This leads to consider a more sophisticated minimisation problem for which the solution is still a local weighted average (but with more complicated weights).

2. Functional locally modelled regression

2.1. *Building of the estimator*

One considers the following minimisation problem ($\mathcal{P}_2$):

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\,\beta(\mathcal{X}_i, \chi))^2 K(h^{-1}|\delta(\mathcal{X}_i, \chi)|),$$

where $\beta(\cdots)$ is a known operator from $\mathcal{H} \times \mathcal{H}$ into $\mathbb{R}$ such that, $\forall \xi \in \mathcal{H},\ \beta(\xi, \xi) = 0$. The locally modelled regression estimator m_{LMRE} of m is the solution for a of the Problem ($\mathcal{P}_2$) and we have

$$m_{\text{LMRE}}(\chi) = {}^{\text{t}}\boldsymbol{u}_1 ({}^{\text{t}}\boldsymbol{Q}_\beta \boldsymbol{K} \boldsymbol{Q}_\beta)^{-1}\, {}^{\text{t}}\boldsymbol{Q}_\beta \boldsymbol{K} \boldsymbol{Y},$$

where

$${}^{\text{t}}\boldsymbol{Q}_\beta = \begin{bmatrix} 1 & \cdots & 1 \\ \beta(\mathcal{X}_1, \chi) & \cdots & \beta(\mathcal{X}_n, \chi) \end{bmatrix},$$

$\boldsymbol{K} = \text{diag}(K(h^{-1}|\delta(\mathcal{X}_1, \chi)|), \ldots, K(h^{-1}|\delta(\mathcal{X}_n, \chi)|))$, $\boldsymbol{Y} = {}^{\text{t}}[Y_1, \ldots, Y_n]$ and $\boldsymbol{u}_1 = {}^{\text{t}}[1, 0] \in \mathbb{R}^2$ (where $^{\text{t}}$ is the transpose symbol). One of the main advantages of this method is that one gets the explicit solution:

$$m_{\text{LMRE}}(\chi) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} Y_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}},$$

with

$$w_{ij} = \beta_i(\beta_i - \beta_j) K_i K_j,$$

where $K_i = K(h^{-1}|\delta(\mathcal{X}_i, \chi)|)$ and $\beta_i = \beta(\mathcal{X}_i, \chi)$. This kernel-type estimator appears as a local weighted average since it can be rewritten as follows:

$$m_{\text{LMRE}}(\chi) = \sum_{j=1}^{n} W_j Y_j, \text{ with } \forall j = 1, \ldots, n, \qquad W_j = \frac{\sum_{i=1}^{n} w_{ij}}{\sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij}}.$$

This explicit expression of $m_{\text{LMRE}}(\chi)$ is interesting for fast computational issues and also allows us to state asymptotic properties. Before going on, it is worth discussing the role of the various parameters introduced in the estimator.

2.2. *Interpretation and comments*

From a statistical point of view, our LMRE-approach assumes that $a + b\,\beta(\cdots\chi)$ is a good approximation of $m(\cdot)$ around χ. Because $\beta(\chi, \chi) = 0$, a is a good approximation of $m(\chi)$ and hence, m_{LMRE} should be a good estimator of m. In this sense, m_{LMRE} is really a functional locally non-constant regression estimator inspired by the local linear nonparametric approaches. The idea of local linear regression has been developed by Muller (1989) and Fan and Gijbels (1992) and the minimax properties of the corresponding smoother have been stated by Fan (1993). Multivariate extension can be found in Ruppert and Wand (1994) and more recently in Lafferty and Wasserman (2008). A good overview on this topic can be found in Wand and Jones (1995 and Fan and Gijbels (1996), and we refer to other authors (Cheng and Hall 2002; Hengartner, Wegkamp, and Matzner-Løber 2002; García-Soidán, González-Manteiga, and Febrero-Bande 2003) for more recent advances.

Clearly, the behaviour of the estimate depends on $\delta(\cdots)$ and $\beta(\cdots)$ whose choices will be crucial. To fix ideas, one can give particular shapes for $\delta(\cdots)$ and $\beta(\cdots)$. For instance, if the functional data are 'smooth' curves, one can try to use the following family of locating functions:

$$\mathrm{loc}_a^{(q)}(\chi_1, \chi_2) = \int \theta(t)(\chi_1^{(q)}(t) - \chi_2^{(q)}(t))\,\mathrm{d}t =< \theta, \quad \chi_1^{(q)} - \chi_2^{(q)} >_{\mathcal{H}},$$

where θ is a given function which can be adapted to the data (see Section 3) ($\chi^{(q)}$ denoting the qth derivative of χ). Choosing $\beta(\cdots)$ in such a family is motivated by the fact that it corresponds to the minimisation problem

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\ < \theta, \mathcal{X}_i^{(q)} - \chi^{(q)} >_{\mathcal{H}})^2 K\left(h^{-1}|\delta(\mathcal{X}_i, \chi)|\right),$$

which can be viewed as a kind of 'local linear' regression approach when one considers a functional explanatory variable. Of course metrics, or more generally semi-metrics also based on derivatives, could be good candidates for locating one curve with respect to another one. For instance, one can define

$$\mathrm{loc}_b^{(q)}(\chi_1, \chi_2) = \left(\int \left(\chi_1^{(q)}(t) - \chi_2^{(q)}(t)\right)^2 \mathrm{d}t\right)^{1/2},$$

which is a semi-metric. This second family of locating functions is particularly well adapted for $\delta(\cdots)$, which measures the proximity between two elements of $\mathcal{H}$.

Finally, lots of locating functions can be defined. Indeed, one can build semi-metrics based on the functional principal components analysis (Dauxois, Pousse, and Romain 1982) or inspired by the partial least squares method (see, for instance, Martens and Naes (1989)). However, in order to make our purpose clear we just introduce here the locating functions used later on for studying the spectrometric functional data set.

As $\delta(\cdots)$, $\beta(\cdots)$ locates also one element of $\mathcal{H}$ with respect to another one which means that $\beta(\cdots)$ can take also negative values. In other words, the locating functions $\delta(\cdots)$ and $\beta(\cdots)$ allow to compute a 'signed' proximity between two curves. Let us remark that the theory allows to take $\beta(\cdots) = \delta(\cdots)$. But, β and δ do not play similar roles. Indeed, β refers to the local behaviour of the regression whereas δ concerns the local weighting. Therefore, as we will see in Section 3, authorising two ways for locating one element with respect to another one (i.e. $\beta \neq \delta$) can allow us to better fit the data from a practical point of view. In that sense, considering $\beta \neq \delta$ can lead us to a more adaptative (or flexible) method.

2.3. *An intermediate flexible local regression*

m_{LMRE} is a trade-off between m_{LCRE} and m_{LLRE}, the latter being the functional local linear regression estimator introduced in a more general way in Barrientos-Marin (2007, pp. 69–71), investigated recently (see Boj et al. (2008) and Baíllo and Grané (2009) for independent observations, and Aneiros-Pérez et al. (2008) in the time-series setting), and defined as the solution of the following functional minimisation problem ($\mathcal{P}_3$):

$$\min_{(a,\psi)\in\mathbb{R}\times\mathcal{H}} \sum_{i=1}^{n} (Y_i - a - < \psi, \mathcal{X}_i - \chi >_{\mathcal{H}})^2 K(h^{-1}|\delta(\mathcal{X}_i, \chi)|),$$

where $< \cdots >_{\mathcal{H}}$ is some inner product defined onto $\mathcal{H}$. Clearly, if one replaces $< \psi, \mathcal{X}_i - \chi >_{\mathcal{H}}$ with $b\beta(\mathcal{X}_i, \chi)$, one gets exactly ($\mathcal{P}_2$), which leads us to the functional locally modelled regression estimator. So, instead of considering the functional parameter ψ (needed for defining m_{LLRE}), one deals with the real parameter b and the fixed operator $\beta(\cdots)$. In this sense, m_{LMRE} is more flexible than m_{LCRE} but less flexible than m_{LLRE}; m_{LMRE} is really a compromise between m_{LCRE} and m_{LLRE}. Moreover, one can enlarge the flexibility of m_{LMRE}. For instance, one can select the operator $\beta(\cdots)$ over some class of operators throughout a cross-validation procedure. Finally, as we will see later on (Section 3.4), the lack of flexibility of m_{LMRE} can be balanced by combining it with the local constant one (m_{LCRE}).

3. LMRE regression in action

The aim of this section is to give an idea of the behaviour of the LMRE method from a practical point of view. To do that, we have chosen some standard curves' data sets coming from chemometrics (Section 3.1). Indeed, we had a choice between many curves' data sets but the following one has been selected for pedagogical reasons, since it is one of the most popular functional data sets commonly used in the functional statistician community to check the feasability and the performances of any new method. The computational issues linked with the practical implementation of the functional locally modelled approach developed in this paper are discussed in 3.2, and the results obtained on the chemometrical data are reported in Section 3.4. One also compares our local regression method with the functional local linear one.

3.1. *A real curves' data set example*

One considers the functional data set $(\boldsymbol{\mathcal{X}}_i, Y_i)_{i=1,\ldots,215}$ where the vector $\boldsymbol{\mathcal{X}}_i = {}^t(\mathcal{X}_i(\lambda_1), \ldots, \mathcal{X}_i(\lambda_{100}))$ is the ith discretised spectrometric curve presented in the introduction and Y_i its corresponding percentage of moisture.

Initially studied by Borggaard and Thodberg (1992), the following data set comes from a quality control problem in the food industry and can be found at http://lib.stat.cmu.edu/datasets/tecator. For $n = 215$ pieces of finely chopped meat, one observes a spectrometric curve by means of a Tecator Infractec Food and Feed Analyser working in the wavelength range of 850–1050 nm by the near-infrared transmission principle (Figure 1), and for each piece of meat, one also knows the percentage of moisture (by means of some chemical process). The aim is clear: can we predict the percentage of moisture in a piece of meat from its spectrometric curve? This is a regression problem involving a functional covariate (namely, the spectrometric curve $\mathcal{X}_i$'s) and a scalar response (namely, the moisture measurement Y_i).

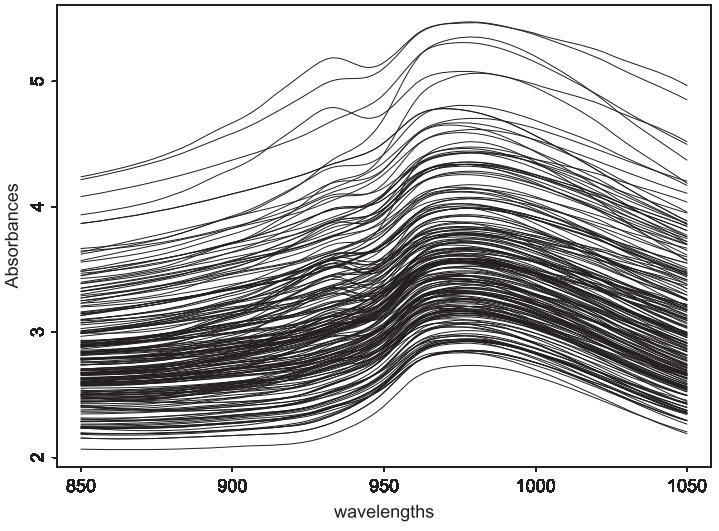

Figure 1. The spectrometric curves.

3.2. *Choosing the bandwidth and the locating functions*

A first important practical aspect consists of fixing the local shape of the regression, that is, the operator $\beta(\cdot,\cdot)$. Because of the smoothness of the spectrometric curves and according to the experience of chemometricians, one proposes to use the family of locating functions $\{\mathrm{loc}_a^{(q)}\}_{q=0,1,2,\ldots}$ defined in Section 2. In addition, the real-valued function θ is selected among the eigenfunctions of the covariance operator $\mathbb{E}\left((\mathcal{X}^{(q)} - \mathbb{E}\mathcal{X}^{(q)}) < \mathcal{X}^{(q)}, \cdot >_{\mathcal{H}}\right)$. From a practical point of view, one gets a discretised version of these eigenfunctions by computing the eigenvectors of the empirical covariance operator $1/|\mathcal{L}| \sum_{i\in\mathcal{L}} (\boldsymbol{\mathcal{X}}_i^{(q)} - \overline{\boldsymbol{\mathcal{X}}}^{(q)})\, {}^t(\boldsymbol{\mathcal{X}}_i^{(q)} - \overline{\boldsymbol{\mathcal{X}}}^{(q)})$, where $\mathcal{L}$ is a subset of $\{1,\ldots,215\}$ ($\mathcal{L}$ will be called learning sample). The idea of choosing θ through the spectral analysis of the covariance operator is driven by the functional principal component analysis methodology (see, for instance, Dauxois et al. (1982); Castro, Lawton, and Sylvestre (1986); Ramsay and Dalzell (1991); Locantore et al. (1999); and Boente and Fraiman (2000) for both practical and theoretical aspects). Indeed, this is a useful tool for exhibing functional directions which can reveal pertinent information. A second crucial point concerns the choice for the locating operator $\delta(\cdot,\cdot)$. As previously, a well-adapted tool is the family of semi-metrics based on the derivatives (i.e. $\mathrm{loc}_b^{(q)}$).

The last point concerns the selection of the smoothing parameter (i.e. the bandwidth h), which is a standard problem in nonparametric statistics. To this end, one uses a cross-validation procedure over kNN-type bandwidths allowing the choice of a local bandwidth (i.e. $h = h(\chi)$ in the definition of $m_{\mathrm{LMRE}}(\chi)$). More precisely, let $h_k(\chi)$ be a local bandwidth for which there are exactly k curves $\mathcal{X}_{i_1},\ldots,\mathcal{X}_{i_k}$ such that $|\delta(\mathcal{X}_{i_l},\chi)| < h_k(\chi)$ and let $m_{\mathrm{LMRE}}^{(-i,k)}(\chi)$ be the kernel-type estimator using $h_k(\chi)$ and computed without the ith unit. Then, one defines an optimal number k_{opt} of neighbours as follows:

$$k_{\mathrm{opt}} = \arg\min_k \sum_i \left(Y_i - m_{\mathrm{LMRE}}^{(-i,k)}(\mathcal{X}_i)\right)^2.$$

This procedure allows us to evaluate m_{LMRE} at any fixed curve χ by using the optimal local bandwidth $h(\chi) := h_{k_{\mathrm{opt}}}(\chi)$.

3.3. *Presentation of the study*

In order to illustrate the pertinence of the LMRE method, the initial sample is randomly split into two subsamples. The first one, $(\mathcal{X}_i, Y_i)_{i\in\mathcal{L}}$, usually called the 'learning sample', allows to build the estimator (i.e. $m^{\mathcal{L}}_{\mathrm{LMRE}}(\cdot)$ with the optimal number of neighbours k_{opt}). The second subsample, $(\mathcal{X}_i, Y_i)_{i\in\mathcal{T}}$, usually called the 'testing sample', allows to predict percentages of moisture (i.e. $\{m^{\mathcal{L}}_{\mathrm{LMRE}}(\mathcal{X}_i)\}_{i\in\mathcal{T}}$). Of course, $\mathcal{L}$ and $\mathcal{T}$ are built such that $\mathcal{L}\cup\mathcal{T} = \{1,\ldots,215\}$ and $\mathcal{L}\cap\mathcal{T} = \phi$. Now, one way to evaluate the performance of this method is to compute the mean square errors (MSE) of prediction:

$$\mathrm{MSE}(\mathcal{L},\mathcal{T}) \;=\; \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}(Y_i - m^{\mathcal{L}}_{\mathrm{LMRE}}(\mathcal{X}_i))^2,$$

which is a usual index for evaluating the quality of fitting to the data. In order to get more robust results, one randomly builds 100 learning and testing samples

$$(\mathcal{L}^{(s)}, \mathcal{T}^{(s)})_{s=1,\ldots,100},$$

which allows us to get the 100 quantities $\mathrm{MSE}(\mathcal{L}^{(s)}, \mathcal{T}^{(s)})$, for $s = 1,\ldots,100$. Moreover, in our study, we always take $|\mathcal{L}| = 165$ (and hence $|\mathcal{T}| = 50$). To show the good behaviour of the LMRE method, we also implemented the functional local constant regression estimate m_{LCRE} (Section 2) with the same bandwidth choice procedure. At last, to give an idea on the predictive performance of our method, one compares it with the functional local linear method (i.e. m_{LLRE}) using a B-splines expansion for the functional parameter ψ and also the same family of semi-metrics ($\{\mathrm{loc}^{(q)}_b\}_{q=0,1,2,\ldots}$).

3.4. *Results and comments*

Here, for the local behaviour $\beta(\cdot,\cdot)$, we have tried several parameters q and θ, and the best results in terms of prediction are obtained for $q = 1$ and for the third eigenfunction (i.e. the eigenfunction associated to the third largest eigenvalue). Concerning the weighting locating function, the best results are obtained for the second derivative (i.e. $\delta(\cdot,\cdot) = \mathrm{loc}^{(2)}_b(\cdot,\cdot)$), and the fixed asymmetrical kernel function is the quadratic one (i.e. $K(u) = (3/2)(1-u^2)1_{[0,1]}$). This confirms the idea that allows an index of proximity δ to be different to the local shape β and makes this LMRE method more flexible in the sense that the fit is better.

Figure 2 displays various *box-and-whiskers* which summarise the distribution of MSE computed over 100 experiments: the left one corresponds to the LCRE regression, the middle one gives an idea of the performance of the LMRE method, and the right one corresponds to the 'AVERAGE' method, which predicts the moisture by using $0.5 * (m_{\mathrm{LCRE}} + m_{\mathrm{LMRE}})$. In this situation, the locally modelled approach seems to globally fit the data better than the local constant one. Actually, looking at the 'AVERAGE' method we see that, if we run both functional nonparametric approaches, we improve the quality of fitting in terms of MSE. This means that for some units, when m_{LCRE} overestimates, m_{LMRE} balances the prediction by underestimating and similarly in reverse. In other words, the LMRE method has to be seen as a complementary nonparametric tool of prediction, and not only as a competitive one with respect to existing functional nonparametric methods. Now, if one compares m_{LMRE} with the functional local linear regression m_{LLRE}, this last one is the best from a predictive point of view with $\delta(\cdot,\cdot) = \mathrm{loc}^{(2)}_b(\cdot,\cdot)$ (which leads to the best predictions). As explained in (Aneiros-Pérez et al. 2008; Boj et al. 2008), $\delta(\cdot,\cdot)$ plays a major role also in the functional local linear approach (and considering only the standard L^2-norm could lead to bad predictions). This good behaviour of the local linear regression is not surprising since m_{LLRE} is a more flexible (i.e. adaptive) method than m_{LMRE}. But, if one looks at

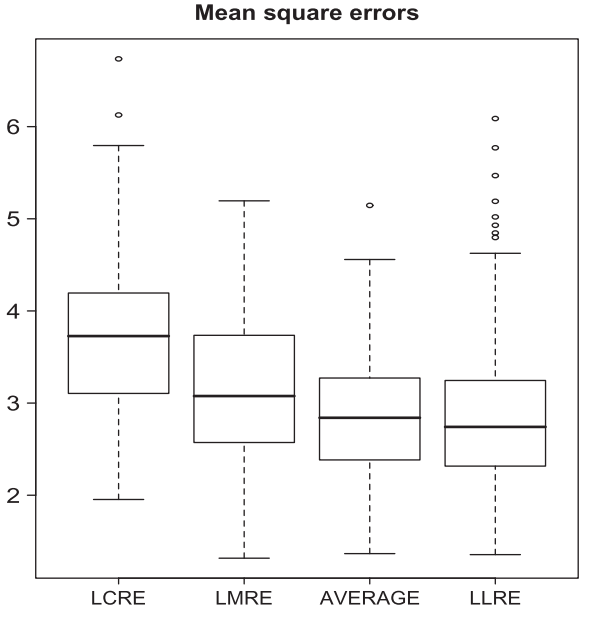

Figure 2. Distribution of the MSE's for each method.

the combination of m_{LCRE} with m_{LMRE} ('AVERAGE'), the predictive performance is very similar to the one reached by the functional local linear regression m_{LLRE}. Moreover, it is worth noting that the computational cost for the method 'AVERAGE' is much smaller than the one needed for implementing m_{LLRE}. To give an idea, if one selects only the bandwidth by cross-validation (the semi-metrics being fixed), the AVERAGE method takes about $6''$ whereas m_{LLRE} needs about $1'$ (10 times more). One can imagine easily that the computational gain is much more important if one implements a cross-validation procedure for choosing semi-metrics. This is also the case when one processes larger data sets.

4. Asymptotic properties

Our goal is to study the asymptotic behaviour of $m_{\mathrm{LMRE}}(\chi)$, the locally modelled estimator of the regression operator $\mathbb{E}(Y|\mathcal{X} = \chi) = m(\chi)$, χ being a fixed element of $\mathcal{H}$. The results are stated in terms of the almost complete convergence which implies the almost sure convergence. Before giving the theorems, let us introduce some assumptions and terminologies. First of all, let us start with a crucial hypothesis concerning the distribution of the functional r.v. $\mathcal{X}$.

(H1) $\varphi_\chi(u_1, u_2) := P(u_1 \leq \delta(\mathcal{X}, \chi) \leq u_2)$, and $\forall u > 0$, $\varphi_\chi(u) := \varphi_\chi(0, u) > 0$.

It is easy to see that

$$\varphi_\chi(u) = P(\mathcal{X} \in B(\chi, u)),$$

where

$$B(\chi, u) = \{\chi' \in \mathcal{H},\ |\delta(\chi, \chi')| \leq u\}.$$

As soon as $|\delta(\cdot, \cdot)|$ defines a metric or, more generally, a semi-metric, $\varphi_\chi(u)$ can be interpreted as the probability of a ball of $\mathcal{H}$ centered at χ and of radius u. When u becomes small, the terminology 'small ball probabilities' is commonly used, which is a field of the probability theory intensively

investigated (see Li and Shao (2001) for an overview on this topic in relation to gaussian processes). Actually, the function φ_χ (which is a direct extension of the small ball probability notion) plays the same role in the functional case as the density function in the finite-dimensional setting. Indeed, it is usual in the multivariate nonparametric setting to estimate a quantity at a point for which one has all around a number of observations large enough. A common way for assuming that is to assume that the density function valued at this point is strictly positive. In the infinite-dimensional setting, there is no reference measure such as the Lebesgue one in the finite-dimensional context, and one has to make a similar assumption without the density notion. This is the goal of Hypothesis (**H1**) which translates in the functional context the fact that we have enough observations around χ and hence it makes sense to estimate the regression operator at this point χ.

As it is standard in nonparametric modelling, one has to consider regularity assumptions. The first ones concern the unknown regression operator m which will be supposed to verify one of the following constraints:

(**H2**$_C$) $m \in \left\{ f : \mathcal{H} \to \mathbb{R},\ \lim_{|\delta(\chi,\chi')|\to 0} f\left(\chi'\right) = f(\chi) \right\}$,

or

(**H2**$_L$) $m \in \left\{ f : \mathcal{H} \to \mathbb{R},\ ,\ \forall \chi' \in \mathcal{H},\ \left| f\left(\chi'\right) - f(\chi) \right| < C|\delta(\chi, \chi')|^\nu \right\}$, where C and ν are fixed in $\mathbb{R}^+$.

Clearly, the first hypothesis is a continuity-type constraint which will allow us to get pointwise convergence. Moreover, as soon as one wishes to state the rate of convergence, one has to introduce more restrictive constraints, which is the role played by the second Lipschitz-type hypothesis.

Another regularity-type constraint is also necessary in order to control the shape of the local functional object β:

(**H3**) $\exists 0 < M_1 < M_2,\ \forall \chi' \in \mathcal{H},\ M_1\,|\delta(\chi, \chi')| \le |\beta(\chi, \chi')| \le M_2\,|\delta(\chi, \chi')|$.

Now, we focus on assumptions concerning the kernel estimator m_{LMRE}. Let us first introduce assumptions on the kernel function K:

(**H4**) K is a function from $\mathbb{R}$ into $\mathbb{R}_+$ differentiable on its support $[0, 1]$.

These kinds of kernels contain the standard asymmetrical kernels used in the literature (uniform, triangle, quadratic, Epanechnikov, etc.). Once this class of kernels are defined, one can propose additional hypotheses acting on the estimator (and also on the distribution of the functional r.v. $\mathcal{X}$):

(**H5**) h is a positive sequence such that $\lim_{n\to\infty} h = 0$, and $\lim_{n\to\infty} \log n/n\, \varphi_\chi(h) = 0$.

(**H6**) $\exists n_0,\ \forall n > n_0,\ 1/\varphi_\chi(h) \int_0^1 \varphi_\chi(zh, h)\, \mathrm{d}/\mathrm{dz}(z^2 K(z))\mathrm{d}z > C > 0$.

(**H7**) $h \int_{B(\chi,h)} \beta(u, \chi)\, \mathrm{d}P(u) = o\left(\int_{B(\chi,h)} \beta(u, \chi)^2\, \mathrm{d}P(u) \right)$.

One ends the listing of hypotheses by focusing on the scalar response Y through its successive conditional moments:

(**H8**) $\forall k = 2, 3, \ldots, \sigma_k : \chi \mapsto \mathbb{E}\left(Y^k | \mathcal{X} = \chi\right)$ is a continuous operator.

Hypotheses (**H2**)–(**H5**) and (**H8**) are standard in the nonparametric functional regression setting and extend what is usually assumed in the classical p-dimensional nonparametric literature (Ferraty and Vieu 2006). Hypothesis (**H6**) precises the behaviour of the bandwidth h in relation with the small ball probabilities and the kernel function K. This hypothesis is not restrictive because one can see easily that fractal processes satisfy it for a wide class of kernel functions (see Appendix 1 for more details). The key new hypothesis is (**H7**) about the local behaviour of

the operator β which models the local shape of the regression. For instance, in the special case where $\beta = \delta$, this assumption means that the local expectation of β is small enough with respect to its moment of second order. If, in addition, the real r.v. $\beta(\mathcal{X}, \chi)$ admits a differentiable density (around 0) with respect to the Lebesgue measure then (**H7**) is satisfied.

Now, we are ready for giving the two main results. The first one states the pointwise almost complete convergence, whereas the second one precises the rate of convergence.

THEOREM 4.1 *Under* (**H1**), (**H2**$_C$), (**H3**)–(**H8**), *we have*

$$m_{\mathrm{LMRE}}(\chi) - m(\chi) \;=\; o_{\mathrm{a.co.}}(1).$$

THEOREM 4.2 *Under* (**H1**), (**H2**$_L$), (**H3**)–(**H8**), *we have*

$$m_{\mathrm{LMRE}}(\chi) - m(\chi) \;=\; O\left(h^{\nu}\right) + O_{\mathrm{a.co.}}\left(\sqrt{\frac{\log n}{n\,\varphi_{\chi}\left(h\right)}}\right).$$

The remainder of this section gives a sketch of the proof of both theorems (details are postponed in Appendices 2 and 3). Let us first introduce the following quantities:

$$m_l(\chi) \;=\; \frac{1}{n(n-1)\,\mathbb{E}w_{12}} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} Y_j^l, \qquad \text{for } l = 0, 1,$$

in such a way that we have $m_{\mathrm{LMRE}}(\chi) = m_1(\chi)/m_0(\chi)$ with $\mathbb{E}m_0(\chi) = 1$. The proof of the above-mentioned theorems is based on the decomposition:

$$\begin{aligned} m_{\mathrm{LMRE}}(\chi) - m(\chi) = \frac{1}{m_0(\chi)}\left[\left(m_1(\chi) - \mathbb{E}m_1(\chi)\right) - \left(m(\chi) - \mathbb{E}m_1(\chi)\right)\right] \\ - \frac{m(\chi)}{m_0(\chi)}\left(m_0(\chi) - 1\right), \end{aligned}$$

and on the following lemmas:

LEMMA 4.3 *Assume that hypotheses* (**H1**), (**H3**)–(**H5**) *are satisfied.*

(i) *If* (**H2**$_C$) *holds, one gets:*

$$m(\chi) - \mathbb{E}m_1(\chi) = o(1),$$

(ii) *If* (**H2**$_L$) *holds, one gets*

$$m(\chi) - \mathbb{E}m_1(\chi) = O(h^{\nu}).$$

LEMMA 4.4 *Suppose that assumptions* (**H1**), (**H2**$_C$), (**H3**)–(**H7**) *are satisfied.*

(i) *One has*

$$m_0(\chi) - 1 = O_{\mathrm{a.co.}}\left(\sqrt{\frac{\log n}{n\,\varphi_{\chi}(h)}}\right),$$

(ii) *In addition, if* (**H8**) *holds, one gets*

$$m_1(\chi) - \mathbb{E}m_1(\chi) = O_{\mathrm{a.co.}}\left(\sqrt{\frac{\log n}{n\,\varphi_{\chi}(h)}}\right).$$

Lemmas 4.3-(i) and 4.4 lead to the statement of Theorem 4.1. Lemma 4.3-(ii) in addition with Lemma 4.4 allows us to get Theorem 4.2. Detailed proofs of these lemmas can be found in the appendix section.

5. Perspectives and conclusion

One can see the LMRE method as a good alternative to the LCRE approach. This good behaviour is observed both for asymptotic point of view and for the spectrometric application, making this functional locally modelled regression very attractive. As emphasised in the implementation, LMRE is also a complementary functional prediction method in the sense that combined with the LCRE, the obtained MSE overpasses those coming from each of LCRE or LMRE. This combination of LCRE and LMRE reaches similar predictive performances than the ones obtained with the functional local linear approach but with a much lower computational cost. That makes this new methodology very attractive for practitioners.

In addition, this work offers very interesting perspectives of investigations. A first direction concerns the statement of theoretical properties with respect to the bandwidth choice (see Benhenni, Ferraty, Rachdi, and Vieu (2007) for theoretical properties of a cross-validated bandwidth in the functional local constant regression or Antoniadis, Paparoditis, and Sapatinas (in press) for an alternative selection procedure based on a risk minimisation criterion). A second track would consist of looking for a more pertinent functional direction (i.e. θ) by proposing an 'optimal' linear combination of eigenfunctions. Of course, this deserves deeper investigation, from both practical and theoretical points of view.

Finally, this work is a step towards local statistical models taking into account functional variables. This is also an encouragement to pursue further investigations in this topic.

Acknowledgements

The authors acknowledges both reviewers for their helpful comments allowing us to improve significantly this work. One of the authors especially thanks the reviewer, who provided all materials for implementing the alternative functional local linear regression (i.e. m_{LLRE}). This strongly contributed to the improvement of our work. The authors wish to express their gratitude to all the members of the STAPH group in Toulouse for their continuous and fruitful support. The whole activity of this group in various aspects of infinitely dimensional statistics are available on line at *http://www.math.univ-toulouse.fr/˜staph/*. It should be noted that most of the results in this paper come from the PhD dissertation of the first author (Barrientos-Marin 2007).

References

Abraham, C., Cornillon, P., Matzner-Löber, E., and Molinari, E. (2003), 'Unsupervised Curve Clustering Using B-Splines', *Scandinavian Journal of Statistics*, 30, 581–595.

Abramovitch, F., and Angelini, C. (2006), 'Testing in Mixed-Effects FANOVA Models', *Journal of Statistical Planning and Inference*, 136, 4326–4348.

Abramovitch, F., Antoniadis, A., Sapatinas, T., and Vidakovic, B. (2004), 'Optimal Testing in a Fixed-Effects Functional Analysis of Variance Model', *International Journal of Wavelets, Multiresolution and Information Processing*, 2, 323–349.

Aneiros-Pérez, G., Cao, R., and Vilar-Fernández, J.M. (2008), 'Functional Methods for Time Series Prediction: A Nonparametric Approach'. Proceedings of IASC2008, Pacifico, Yokohama, Japan, pp. 91–100.

Antoniadis, A., and Sapatinas, T. (2007), 'Estimation and Inference in Functional Mixed-Effects Models', *Computational Statistics & Data Analysis*, 51, 4793–4813.

Antoniadis, A., Paparoditis, E., and Sapatinas, T. (in press), 'Bandwidth Selection for Functional Time Series Prediction', *Statistics & Probability Letters*.

Baíllo, A., and Grané, A. (2009), 'Functional Local Linear Regression with Functional Predictor and Scalar Response', *Journal of Multivariate Analysis*, 100, 102–111.

Barrientos-Marin, J. (2007), 'Some Practical Problems of Recent Nonparametric Procedures: Testing, Estimation, and Application', Ph.D. thesis, Universidad de Alicante, Departamento de Fundamentos del Análisis Economico, Spain.

Benhenni, K., Ferraty, F., Rachdi, M., and Vieu, P. (2007), 'Nonparametric Regression for Functional Data: Optimal Local Bandwidth Choice', *Computational Statistics*, 22, 353–369.
Boente, G., and Fraiman, R. (2000), 'Kernel-Based Functional Principal Components', *Statistics & Probability Letters*, 48, 335–345.
Boj, E., Delicado, P., and Fortiana, J. (2008), 'Local Linear Functional Based on Weighted Distance-Based Regression', in *Functional and Operatorial Statistics*, eds. S. Dabo-Niang and F. Ferraty, Heidelberg: Physica-Verlag, pp. 57–64.
Borggaard, C., and Thodberg, H. (1992), 'Optimal Minimal Neural Interpretation of Spectra', *Analytical Chemistry*, 64, 545–551.
Burba, F., Ferraty, F., and Vieu, P. (2009), 'k-Nearest Neighbor Method in Functional Nonparametric Regression', *Journal of Nonparametric Statistics*, 21, 453–469.
Castro, P.E., Lawton, W.H., and Sylvestre, E.A. (1986), 'Principal Modes of Variation for Processes with Continuous Sample Curves', *Technometrics*, 28, 329–337.
Cheng, M.Y., and Hall, P. (2002), 'Error-Dependent Smoothing Rules in Local Linear Regression', *Statistica Sinica*, 12, 429–447.
Dabo-Niang, S., and Ferraty, F. (2008), *Functional and Operatorial Statistics*, eds. S. dabo-Niang and F. Ferraty, Heidelberg: Physica-Verlag.
Dabo-Niang, S., Ferraty, F., and Vieu, P. (2004), *Nonparametric Unsupervised Classification of Satellite Wave Altimeter Forms*, Proceedings in Computational Statistics, ed. J. Antoch, Heidelberg: Physica-Verlag, pp. 879–886.
Dauxois, J., Pousse, A., and Romain, Y. (1982), 'Asymptotic Theory for the Principal Component Analysis of a Random Vector Function: Some Application to Statistical Inference', *Journal of Multivariate Analysis*, 12, 136–154.
Davidian, M., Lin, X., and Wang, J.L. (2004), 'Introduction to the Emerging Issues in Longitudinal and Functional Data Analysis (with discussion)', *Statistica Sinica*, 14, 613–629.
Fan, J. (1993), 'Local Linear Regression Smoother and Their Minimax Efficiencies', *The Annals of Statistics*, 21, 196–216.
Fan, J., and Gijbels, I. (1992), 'Variable Bandwidth and Local Linear Regression Smoothers', *The Annals of Statistics*, 20, 2008–2036.
Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman & Hall.
Ferraty, F., and Vieu, P. (2002), 'The Functional Nonparametric Model and Application to Spectrometric Data', *Computational Statistics*, 17, 545–564.
Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, New York: Springer.
Frank, I.E., and Friedman, J.H. (1993), 'A Statistical View of Some Chemometrics Regression Tools (with discussions)', *Technometrics*, 35, 109–148.
García-Soidán, P., González-Manteiga, W., and Febrero-Bande, M. (2003), 'Local Linear Regression Estimation of the Variogram', *Statistics & Probability Letters*, 64, 169–179.
Gasser, T., Hall, P., and Presnell, B. (1998), 'Nonparametric Estimation of the Mode of a Distribution of Random Curves', *Journal of the Royal Statistical Society, Series B*, 60, 681–691.
González Manteiga, W., and Vieu, P. (2007), 'Introduction to the Special Issue on Statistics for Functional Data', *Computational Statistics & Data Analysis*, 51, 4788–4792.
Hall, P., Poskitt, P., and Presnell, D. (2001), 'A Functional Data-Analytic Approach to Signal Discrimination', *Technometrics*, 35, 140–143.
Hengartner, N., Wegkamp, M., and Matzner-Løber, E. (2002), 'Bandwidth Selection for Local Linear Regression Smoothers', *Journal of the Royal Statistical Society, Series B*, 64, 791–804.
Lafferty, J., and Wasserman, L. (2008), 'Rodeo: Sparse, Greedy Nonparametric Regression', *The Annals of Statistics*, 36, 36–63.
Li, W.V., and Shao, Q.M. (2001), 'Gaussian Processes: Inequalities, Small Ball Probabilities and Applications', in *Stochastic Processes: Theory and Methods*, Handbook of Statistics (Vol. 19), Amsterdam: North-Holland.
Locantore, N., Marron, J.S., Simpson, D., Tripoli, N., Zhang, J., and Cohen, K. (1999), 'Robust Principal Component Analysis for Functional Data (with discussion)', *Test*, 8, 1–74.
Martens, H., and Naes, T. (1989), *Multivariate Calibration*, New York: Wiley.
Muller, H.G. (1989), 'Weighted Local Regression and Kernel Methods for Nonparametric Curve', *Journal of the American Statistical Association*, 82, 231–328.
Nadaraya, E.A. (1964), 'On Estimating Regression', *Theory of Probability and its Applications*, 10, 186–196.
Ramsay, J., and Dalzell, C. (1991), 'Some Tools for Functional Data Analysis', *Journal of the Royal Statistical Society, Series B*, 53, 539–572.
Ramsay, J., and Silverman, B.W. (2002), *Applied Functional Data Analysis: Methods and Case Studies*, New York: Springer-Verlag.
Ramsay, J., and Silverman, B.W. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer-Verlag.
Ramsay, J., Altman, N., and Bock, R. (1994), 'Variation in Height Acceleration in the Fels Growth Data', *The Canadian Journal of Statistics*, 22, 89–102.
Ruppert, D., and Wand, M. (1994), 'Multivariate Locally Weighted Least Squares Regression', *The Annals of Statistics*, 22, 1346–1370.
Valderrama, M. (2007), 'Introduction to the Special Issue on Modelling Functional Data in Practice', *Computational Statistics*, 22, 331–334.
Wand, M., and Jones, C. (1995), *Kernel Smoothing*, Monographs on Statistics and Applied Probability (Vol. 60), London: Chapman & Hall.
Watson, G.S. (1964), 'Smooth Regression Analysis', *Sankhya Series A*, 26, 359–372.

Appendix 1. Remark on (H6)

Let us investigate here the special case where the functional variable $\mathcal{X}$ is a fractal process of order δ ($\delta > 0$) such that

$$\lim_{\epsilon \to 0} \sup_{t \in [0,1]} \left| \frac{\varphi_\chi(t\epsilon)}{t^\delta \epsilon^\delta} - C_\chi \right| = 0,$$

where C_χ is a constant which does not depend on t and ϵ. This implies that, for any ϵ small enough, $\varphi_\chi(\epsilon) \sim C_\chi \epsilon^\delta$. Then, it is easy to state

$$\int_0^1 \varphi_\chi(uh, h) \frac{\mathrm{d}}{\mathrm{d}u}(u^2 K(u))\,\mathrm{d}u \;=\; C_\chi h^\delta \int_0^1 (1-u^\delta) \frac{\mathrm{d}}{\mathrm{d}u}(u^2 K(u))\,\mathrm{d}u + o(h^\delta).$$

Now, one considers the family of kernels indexed by $\alpha > 0$ and defined by $K_\alpha(u) = (\alpha + 1/\alpha)(1-u^\alpha)1_{[0,1]}(u)$. This family of kernels contain standard asymmetrical ones (triangle, quadratic). It comes with trivial calculus that

$$\int_0^1 \varphi_\chi(uh, h) \frac{\mathrm{d}}{\mathrm{d}u}(u^2 K(u))\,\mathrm{d}u \;=\; \frac{(\alpha+1)\delta}{(\delta+2)(\alpha+\delta+2)} C_\chi h^\delta + o(h^\delta),$$

which leads us to assumption (**H6**) as soon as h is small enough (i.e. as soon as n is large enough). In the same way, (**H6**) holds when one considers the uniform kernel $1_{[0,1]}(\cdots)$.

In order to complete this section, one can mention that (**H6**) is satisfied for a much wider class of functional random variables (i.e. Hilbertian squared integrable ones) as soon as one considers suitable semi-metrics δ (for more details, see Lemma 13.6, (Ferraty and Vieu 2006, Lemma 13.6, p. 213)).

Appendix 2. Preliminary technical lemma

In the following C is a generic constant ($0 < C < \infty$).

LEMMA A.1 *Under* (**H1**), (**H3**)–(**H6**), *we have*

(i) $\forall (k,l) \in \mathbb{N}^* \times \mathbb{N}$, $\mathbb{E}\left(K_1^k\,|\beta_1|^l\right) \le C\,h^l\,\varphi_\chi(h)$,
(ii) $\mathbb{E}\left(K_1\,\beta_1^2\right) > C\,h^2\,\varphi_\chi(h)$.

Proof

(i) One starts by using (**H3**), which implies

$$K_1^k\,|\beta_1|^l\,h^{-l} \le C\,K_1^k\,|\delta(\mathcal{X}_1,\chi)|^l\,h^{-l},$$

and because the kernel K is bounded on $[0,1]$, one gets

$$K_1^k\,|\beta_1|^l\,h^{-l} \le C\,|\delta(\mathcal{X}_1,\chi)|^l\,h^{-l} 1_{[0,1]}\left(h^{-1}\,|\delta(\mathcal{X}_1,\chi)|\right),$$

and thus, we have

$$\mathbb{E}\left(K_1^k\,|\beta_1|^l\,h^{-l}\right) \le C\,\varphi_\chi(h),$$

which is the claimed result.

(ii) By using (**H3**), it is easy to see that

$$\mathbb{E}K_1\,\beta_1^2 \;>\; C\,\mathbb{E}\delta(\mathcal{X}_1,\chi)^2\,K_1.$$

Moreover, one can write

$$\begin{aligned}
\mathbb{E}\left(K_1\,\frac{\delta(\chi,\mathcal{X}_1)^2}{h^2}\right) &= \int_0^1 t^2 K(t)\,\mathrm{d}P^{|\delta(\chi,\mathcal{X})|/h}(t), \\
&= \int_0^1 \left(\int_0^t \left(\frac{\mathrm{d}}{\mathrm{d}u}(u^2K(u))\right)\mathrm{d}u\right)\mathrm{d}P^{|\delta(\chi,\mathcal{X})|/h}(t), \\
&= \int_0^1 \left(\int_0^1 1_{[u,1]}(t)\,\mathrm{d}P^{|\delta(\chi,\mathcal{X})|/h}(t)\right)\frac{\mathrm{d}}{\mathrm{d}u}(u^2K(u))\mathrm{d}u,
\end{aligned}$$

the last equation coming from the Fubini's theorem. In addition, it is easy to check that

$$\int_0^1 1_{[u,1]}(t)\,\mathrm{d}P^{|\delta(\chi,\mathcal{X})|/h}(t) = P\left(u\,h \le |\delta(\chi,\mathcal{X})| \le h\right),$$

in order to write

$$\mathbb{E}\left(K_1\,\frac{\delta(\chi,\mathcal{X}_1)^2}{h^2}\right) = \int_0^1 \varphi_\chi(uh,h)\,\frac{\mathrm{d}}{\mathrm{d}u}(u^2K(u))\mathrm{d}u.$$

It suffices to use (**H6**) for obtaining the desired lower bound, which ends the proof of Lemma A.1-(ii).

■

Appendix 3. Proofs of main results

Proof of Lemma 4.3 On one side, we have

$$\mathbb{E}m_l(\chi) = \frac{1}{\mathbb{E}w_{12}}\mathbb{E}w_{12}Y_2^l,$$

and on the other side, it holds

$$\mathbb{E}\mathbb{E}\left(m_1(\chi)|\mathcal{X}_2\right) = \frac{1}{\mathbb{E}w_{12}}\mathbb{E}\left(w_{12}\mathbb{E}\left(Y_2|\mathcal{X}_2\right)\right),$$

which allows us to write:

$$\begin{aligned}|m(\chi) - \mathbb{E}m_1(\chi)| &= \frac{1}{|\mathbb{E}w_{12}|}\left|\mathbb{E}\left(w_{12}(m(\chi) - m(\mathcal{X}_2)\right)\right|,\\ &\le \sup_{\chi'\in B(\chi,h)}|m(\chi) - m(\chi')|.\end{aligned}$$

It suffices to consider ($\mathbf{H2}_C$) in order to get Lemma 4.3-(i). However, if one uses ($\mathbf{H2}_L$) instead of ($\mathbf{H2}_C$), it is clear that

$$\sup_{\chi'\in B(\chi,h)}|m(\chi) - m(\chi')| = O(h^v),$$

which leads us to Lemma 4.3-(ii). ■

Proof of Lemma 4.4

(ii) Let us start the proof by remarking that

$$\begin{aligned}m_1(\chi) = \underbrace{\frac{n^2h^2\varphi_\chi(h)^2}{n(n-1)\,\mathbb{E}w_{12}}}_{Q}\Bigg[&\underbrace{\left(\frac{1}{n}\sum_{j=1}^n\frac{K_jY_j}{\varphi_\chi(h)}\right)}_{S_1}\underbrace{\left(\frac{1}{n}\sum_{i=1}^n\frac{K_i\beta_i^2}{h^2\varphi_\chi(h)}\right)}_{S_2}\\ &-\underbrace{\left(\frac{1}{n}\sum_{j=1}^n\frac{K_j\beta_jY_j}{h\varphi_\chi(h)}\right)}_{S_3}\underbrace{\left(\frac{1}{n}\sum_{i=1}^n\frac{K_i\beta_i}{h\varphi_\chi(h)}\right)}_{S_4}\Bigg],\end{aligned} \tag{A1}$$

which allows us to write

$$m_1(\chi) - \mathbb{E}m_1(\chi) \;=\; Q\left[S_1S_2 - \mathbb{E}(S_1S_2) - (S_3S_4 - \mathbb{E}(S_3S_4))\right].$$

At first, one has

$$\begin{aligned}S_1S_2 - \mathbb{E}(S_1S_2) = &(S_1 - \mathbb{E}S_1)(S_2 - \mathbb{E}S_2) + (S_2 - \mathbb{E}S_2)\mathbb{E}S_1\\ &+ (S_1 - \mathbb{E}S_1)\mathbb{E}S_2 + \mathbb{E}S_1\mathbb{E}S_2 - \mathbb{E}S_1S_2.\end{aligned} \tag{A2}$$

It remains to study each term of the decomposition (A2).

(a) Let us write

$$S_1 - \mathbb{E}S_1 \;=\; \frac{1}{n}\sum_{j=1}^{n} \frac{K_j Y_j - \mathbb{E}K_j Y_j}{\varphi_\chi(h)} \;=\; \frac{1}{n}\sum_{j=1}^{n} Z_{1j}.$$

In order to apply an exponential inequality, we focus on the absolute moments of the r.r.v. Z_{1i}:

$$\mathbb{E}|Z_{1i}^m| = \mathbb{E}\left|\varphi_\chi(h)^{-m}(K_jY_j - \mathbb{E}K_jY_j)^m\right|, \tag{A3}$$

$$= \varphi_\chi(h)^{-m}\mathbb{E}\left|\sum_{k=0}^{m} c_{k,m}(K_jY_j)^k(\mathbb{E}K_jY_j)^{m-k}(-1)^{m-k}\right|,$$

$$\leq \varphi_\chi(h)^{-m}\sum_{k=0}^{m} c_{k,m}\mathbb{E}(K_1^k\sigma_k(\mathcal{X}_1))\,|\mathbb{E}K_1\,m(\mathcal{X}_1)|^{m-k}, \tag{A4}$$

the last inequality is obtained by conditioning on $\mathcal{X}_1$. In addition, (**H2**$_C$) implies that $m(\mathcal{X}_1) = m(\chi) + o(1)$ whereas one gets $\sigma_k(\mathcal{X}_1) = \sigma_k(\chi) + o(1)$ as soon as (**H8**) is checked. This, combined with (A4), allows us to write:

$$\mathbb{E}|Z_{1i}^m| = O\left(\varphi_\chi(h)^{-m}\sum_{k=0}^{m}\mathbb{E}(K_1^k)(\mathbb{E}K_1)^{m-k}\right),$$

$$= O\left(\max_{k\in\{0,\ldots,m\}}\varphi_\chi(h)^{-k+1}\right),$$

$$= O\left(\varphi_\chi(h)^{-m+1}\right), \tag{A5}$$

knowing that $\mathbb{E}K_1^k = O(\varphi_\chi(h))$ (see Lemma A.1-(i) with $l = 0$). Finally, it suffices to apply Corollary A.8-(ii) in (Ferraty and Vieu 2006) with $a_n^2 = \varphi_\chi(h)^{-1}$ to get

$$S_1 - \mathbb{E}S_1 \;=\; O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n\,\varphi_\chi(h)}}\right). \tag{A6}$$

(b) In the same way, writing

$$S_2 - \mathbb{E}S_2 \;=\; \frac{1}{n}\sum_{i=1}^{n} \frac{K_i\beta_i^2 - \mathbb{E}(K_1\beta_1^2)}{h^2\,\varphi_\chi(h)} \;=\; \frac{1}{n}\sum_{i=1}^{n} Z_{2i},$$

it easy to see that

$$\mathbb{E}|Z_{2i}^m| \;\leq\; h^{-2m}\varphi_\chi(h)^{-m}\sum_{k=0}^{m} c_{k,m}\mathbb{E}(K_1^k\beta_1^{2k})\left(\mathbb{E}K_1\,\beta_1^2\right)^{m-k},$$

which, combined with Lemma A.8-(i) when $l = 2$ implies that

$$\mathbb{E}|Z_{2i}^m| \;=\; O\left(\varphi_\chi(h)^{-m+1}\right).$$

Once again, one can apply Corollary A.8-(ii) in Ferraty and Vieu (2006) with $a_n^2 = \varphi_\chi(h)^{-1}$ to get

$$S_2 - \mathbb{E}S_2 \;=\; O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n\,\varphi_\chi(h)}}\right). \tag{A7}$$

(c) It is easy to see that

$$\mathbb{E}S_1 = \varphi_\chi(h)^{-1}\mathbb{E}K_1Y_1,$$

$$= \varphi_\chi(h)^{-1}\mathbb{E}K_1m(\mathcal{X}_1),$$

and because $m(\mathcal{X}_1) = m(\chi) + o(1)$, one gets $\mathbb{E}S_1 = O(1)$. Similarly, one can state that $\mathbb{E}S_2 = O(1)$. Now, it remains to study the quantity $\mathbb{E}S_1\mathbb{E}S_2 - \mathbb{E}S_1S_2$. To do that, let us remark that

$$\mathbb{E}S_1\mathbb{E}S_2 - \mathbb{E}S_1S_2 \;=\; \left(1 - \frac{n(n-1)}{n^2}\right)h^{-2}\varphi_\chi(h)^{-2}\mathbb{E}(K_1\beta_1^2)\mathbb{E}(K_1Y_1) + O\left(\left(n\,\varphi_\chi(h)\right)^{-1}\right).$$

Using similar arguments as previously, it is easy to see that

$$\mathbb{E}S_1\mathbb{E}S_2 - \mathbb{E}S_1S_2 \;=\; O\left(\left(n\,\varphi_\chi(h)\right)^{-1}\right),$$

which is negligible with respect to $\sqrt{\log n/(n\,\varphi_\chi(h))}$.

Finally, items (a)–(c) states that

$$S_1 S_2 - \mathbb{E}(S_1 S_2) \;=\; O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n\,\varphi_\chi(h)}}\right). \tag{A8}$$

By similar arguments, one can state

$$S_3 S_4 - \mathbb{E}(S_3 S_4) \;=\; O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n\,\varphi_\chi(h)}}\right). \tag{A9}$$

Now, to achieve the proof, we have to study the upper bound of the quantity Q. Firstly, note that

$$h|\mathbb{E}\beta K_1| \le C\,h\left|\int_{B(\chi,h)} \beta(u,\chi)\,\mathrm{d}P(u)\right|,$$

and (**H7**) implies that

$$h|\mathbb{E}\beta K_1| = o\left(\int_{B(\chi,h)} \beta(u,\chi)^2\,\mathrm{d}P(u)\right).$$

By applying Lemma A.1-(i) with $K = 1_{[0,1]}$, $k = 1$, and $l = 2$, one gets $\int_{B(\chi,h)} \beta(u,\chi)^2\,\mathrm{d}P(u) \le C\,h^2\varphi_\chi(h)$, which implies that

$$\mathbb{E}\beta_1 K_1 \;=\; o\left(h\,\varphi_\chi(h)\right).$$

Now, Lemma A.1-(ii) and the last result allow us to write

$$\mathbb{E}w_{12} = \mathbb{E}(\beta_1^2 K_1)\mathbb{E}K_1 - (\mathbb{E}\beta_1 K_1)^2 \;>\; C\,h^2\,\varphi_\chi(h)^2.$$

As a direct consequence, one has

$$Q \;=\; O(1),$$

which ends the proof of Lemma 4.4-(ii).

(i) This result can be deduced directly from Lemma 4.4-(ii) by taking $Y_i \equiv 1$. In this case, Hypothesis (**H8**) is not necessary.

■