Stat Papers (2011) 52:751–771
DOI 10.1007/s00362-009-0280-2

REGULAR ARTICLE

Automatic estimation procedure in partial linear model with functional data

Germán Aneiros-Pérez · Philippe Vieu

Received: 13 January 2009 / Revised: 16 August 2009 / Published online: 27 October 2009

Abstract Partial linear modelling ideas have recently been adapted to situations when functional data are observed. This paper aims to complete the study of such model by proposing a fully automatic estimation procedure. This is achieved by constructing a data-driven method to choose the smoothing parameters entered in the nonparametric components of the model. The asymptotic optimality of the method is stated and its practical interest is illustrated on finite size Monte Carlo simulated samples.

Keywords Bandwidth selection · Cross-validation · Functional data · Partial linear regression

Mathematics Subject Classification (2000) MSC 62G08 · MSC 62G20

1 Introduction

The partial linear model has been widely used in the literature for multivariate data, as a nice way to balance the trade-off between flexibility and dimensionality of the model. A selected set of references in this field includes the papers by Engle et al. (1986), Robinson (1988) and Speckman (1988) (see also the general monograph by Härdle et al. (2000)). Then, in modern applied statistics, it is more and more usual to have at hand functional datasets, such as curves or images, and it is a challenge

G. Aneiros-Pérez (✉)
Facultad de Informática, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
e-mail: ganeiros@udc.es

P. Vieu
Université Paul Sabatier, Toulouse, France

for the scientific community to develop new statistical models/tools to deal with such infinite dimensional problems. This field has been popularized by the books of Ramsay and Silverman (1997, 2002, 2005). Starting with the book by Ferraty and Vieu (2006) the nonparametric ideas became to be popular in functional regression settings, while the wide recent bibliographical discussion provided by Ferraty and Vieu (2009) emphasizes on the under-development of semi-parametric statistics in this framework. Recently, the partial linear modelling ideas have been adapted to this new kind of data and this can be seen as a first advance in semi-parametric modelling for functional data. More precisely, Aneiros-Pérez and Vieu (2006) stated asymptotic properties under this semi-parametric functional model and showed its interest in an applied case study.

The key question, not resolved until now, concerns the choice of the smoothing factor entered in the nonparametric functional component of the model. As with any nonparametric procedure (functional or not), it is necessary to provide a fully automatic procedure (that is, a data-driven rule) to select this parameter to make the method reliable for a wide scope of users. This is the gap that we aim to bridge in this contribution.

This paper is organized as follows. The model and the estimates are briefly presented in Sect. 2. Then, in Sect. 3, the bandwidth selection problem is addressed using cross-validation ideas. Special attention will be paid to the local adaptive selected bandwidth, given its ability to capture all the local features of the functional data. The asymptotic optimality, for both global and locally adaptive selection rules, is derived in Sect. 4. Section 5 reports a simulation study designed to observe how our bandwidth selectors perform when the sample size is finite, and to compare the global and local procedures. Concluding remarks are given in Sect. 6. Finally, the Appendix presents the technical proofs.

2 The model and estimators

The Semi-Functional Partial Linear Regression (SFPLR) model can be written as:

$$Y = r\left(X_1, \ldots, X_p, T\right) + \varepsilon = \sum_{j=1}^{p} X_j \beta_j + m(T) + \varepsilon, \tag{1}$$

where X_j $(j = 1, \ldots, p)$ are real explanatory variables, T is another explanatory variable but of functional nature, ε is a random error satisfying

$$E\left(\varepsilon \mid X_1, \ldots, X_p, T\right) = 0,$$

β_j $(j = 1, \ldots, p)$ are unknown real parameters and m is an unknown smooth operator. Identifiability of the model is discussed along Remark 1 in Sect. 4. The functional explanatory variable T is valued in some abstract semi-metric space $\mathcal{H}$, and so the target m (respectively r) is a non-linear real valued operator defined on $\mathcal{H}$ (respectively, on $R^p \times \mathcal{H}$, where R denotes de set of real numbers). In the remainder of the paper, we denote by $d(\cdot, \cdot)$ the semi-metric on $\mathcal{H}$ and all the topological notions used derive from the topology $\mathcal{T}_d$ associated with d. This model and the following estimates were first proposed by Aneiros-Pérez and Vieu (2006). The remark in Sect. 4.1 will emphasize on the strong necessity of using such a general abstract semi-metric space, not only for seeking mathematical generality but mainly because it is a key tool for insuring the good behavior of the statistical procedures.

Assume that we have a sample of n independent and identically distributed vectors valued in $R^{p+1} \times \mathcal{C}$ ($\mathcal{C} \subset \mathcal{H}$). These vectors will be denoted from now on by

$$\left\{\left(Y_i, X_{i1}, \ldots, X_{ip}, T_i\right)^T\right\}_{i=1}^n .$$

The SFPLR model can be rewritten by assuming that

$$Y_i = \sum_{j=1}^{p} X_{ij}\beta_j + m(T_i) + \varepsilon_i \quad (i = 1, \ldots, n),$$

where

$$E\left(\varepsilon_i \mid X_{i1}, \ldots, X_{ip}, T_i\right) = 0.$$

As proposed in Aneiros-Pérez and Vieu (2006), we estimate the vector of parameters $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)^T$ and the function m by means of

$$\widehat{\boldsymbol{\beta}}_h = \left(\widetilde{\mathbf{X}}_h^T \widetilde{\mathbf{X}}_h\right)^{-1} \widetilde{\mathbf{X}}_h^T \widetilde{\mathbf{Y}}_h \tag{2}$$

and

$$\widehat{m}_h(t) = \sum_{i=1}^{n} w_{n,h}(t, T_i)(Y_i - \mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_h), \tag{3}$$

respectively. In these estimators, $h > 0$ is a smoothing parameter that typically appears in any setting of nonparametric estimations. Furthermore, we have denoted $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^T$ with $\mathbf{X}_i = \left(X_{i1}, \ldots, X_{ip}\right)^T$, $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and, for any $(n \times q)$-matrix $\mathbf{A}$ ($q \geq 1$), $\widetilde{\mathbf{A}}_h = (\mathbf{I} - \mathbf{W}_h)\mathbf{A}$, where $\mathbf{W}_h = \left(w_{n,h}\left(T_i, T_j\right)\right)_{i,j}$ with $w_{n,h}(\cdot, \cdot)$ being a weight function that can take different forms. Thus, we estimate the regression function $r\left(x_1, \ldots, x_p, t\right)$ by means of $\widehat{r}_h\left(x_1, \ldots, x_p, t\right) = \mathbf{x}^T \widehat{\boldsymbol{\beta}}_h + \widehat{m}_h(t)$, where we have denoted $\mathbf{x} = \left(x_1, \ldots, x_p\right)^T$.

3 Bandwidth selection

3.1 Global cross-validation selection rule

Starting with Härdle and Marron (1985), much literature has been devoted to the construction of automatic bandwidth selection rules in nonparametric kernel modelling.

On the one hand, advances in the specific context of partial linear modelling (but with non-functional data) have been reported by Linton (1995) and Aneiros-Pérez and Quintela-del-Río (2001), while on the other, Rachdi and Vieu (2007) have recently provided some advances in pure nonparametric functional settings. Usually, in any nonparametric smoothing parameter choice problem, most of the selection techniques consist in looking for data-driven values of h minimizing (at least asymptotically) some measure of accuracy of the estimate. In our model, this could be the following estimation error:

$$\text{ASE}\,(h) = n^{-1} \sum_{i=1}^{n} \left(\widehat{r}_h \left(X_{i1}, \ldots, X_{ip}, T_i\right) - r \left(X_{i1}, \ldots, X_{ip}, T_i\right)\right)^2 G\,(T_i),$$

where G is some (known) nonnegative weight function. In standard multivariate situations, cross-validation is a quite popular procedure for attacking the problem but competitive alternatives (such as plug-in rule or bootstrapping) exist. From a technical point of view, cross-validation is easier to use because it only requires knowing the asymptotic rates of convergence of the error, while bootstrapping or plug-in require the constant terms involved in these asymptotic expansions. In the functional setting, the statement of precise asymptotic expansion is generally difficult. In the pure nonparametric functional setting (see for instance Delsol (2009) for the most recent developments), these constants are complicated, while in our partial linear functional context they are still unknown. More discussion on these alternative ways for thinking bandwidth selection problems, with bibliographical references, is included in Ferraty and Vieu (2009).

For all these reasons, it makes sense to construct our selection rule by trying to adapt cross-validation ideas to take into account both specificities of the model: the functional feature of the data and the semi-parametric modelling. The cross-validated bandwidth selector $\widehat{h}$ is defined as

$$\widehat{h} = \arg\min_{h \in H_n} \text{CV}\,(h),$$

where

$$\text{CV}\,(h) = n^{-1} \sum_{i=1}^{n} \left(Y_i - \widehat{r}_h^{\,i} \left(X_{i1}, \ldots, X_{ip}, T_i\right)\right)^2 G\,(T_i),$$

with $\widehat{r}_h^{\,i}\left(x_1, \ldots, x_p, t\right) = \mathbf{x}^T \widehat{\boldsymbol{\beta}}_h + \widehat{m}_h^i(t)$. In this expression, $\widehat{m}_h^i(t)$ denotes the leave-one-out estimator which is constructed like (3) but where, after estimating $\boldsymbol{\beta}$, the i-th data $\left(Y_i, X_{i1}, \ldots, X_{ip}, T_i\right)^T$ is eliminated from the sample. In this paper we will focus on the weights

$$w_{n,h}(t, T_i) = \frac{K\,(d\,(t, T_i)/h)}{\sum_{j=1}^{n} K\left(d\left(t, T_j\right)/h\right)},$$

where K is a function from $[0, \infty)$ into $[0, \infty)$ (these weights are a functional version of the Nadaraya-Watson type weights; see Ferraty and Vieu 2006, Chap. 5). Therefore, the expression of $\widehat{m}_h^i(t)$ is

$$\widehat{m}_h^i(t) = \sum_{j \neq i}^{n} w_{n,h}^i(t, T_j) \left(Y_j - \mathbf{X}_j^T \widehat{\boldsymbol{\beta}}_h \right),$$

where

$$w_{n,h}^i(t, T_j) = \frac{K\left(d\left(t, T_j\right)/h\right)}{\sum_{k \neq i}^{n} K\left(d\left(t, T_k\right)/h\right)}.$$

3.2 Location adaptive selection rule

The selection procedure discussed before is called global in the sense that it is selecting the same parameter for any functional data t at which the estimate is computed. However, various earlier works on one-dimensional problems (see for instance Staniswalis 1989; Vieu 1991) show evidence suggesting that such an approach has the main drawback of not being sufficiently flexible to capture local information in the data. In the recent work by Benhenni et al. (2007), a local adaptive bandwidth improves more in the functional setting than in the univariate, which is expected since the higher complexity of the data (coming from their infinite dimensional feature) would yield more local variation.

One way to construct the location adaptive estimate is to choose, for each t, a specific bandwidth h_t, with the aim of not minimizing the whole errors but only a local version of quadratic errors. In our model, this could be the following local measurement of error

$$\mathrm{ASE}_t(h) = n^{-1} \sum_{i=1}^{n} \left(\widehat{r}_h\left(X_{i1}, \ldots, X_{ip}, T_i\right) - r\left(X_{i1}, \ldots, X_{ip}, T_i\right)\right)^2 G_{n,t}(T_i),$$

where, for each t, the weight function $G_{n,t}$ is chosen to be more and more concentrated around t. Specific details on this local sequence of weight functions will be given later. For now, just think that $G_{n,t}$ could be taken in the form

$$G_{n,t}(t') = f_n(d(t, t')),$$

where f_n is some density function (Gaussian, for instance, but not necessarily) with zero mean and with variance tending to zero as n approaches infinity. Similarly, as the global cross-validation criterion was constructed before, the local selection criterion is defined as

$$\mathrm{CV}_t(h) = n^{-1} \sum_{i=1}^{n} \left(Y_i - \widehat{r}_h^i\left(X_{i1}, \ldots, X_{ip}, T_i\right)\right)^2 G_{n,t}(T_i),$$

and the location adaptive bandwidth is

$$\widehat{h}_t = \arg\min_{h \in H_n} \mathrm{CV}_t(h).$$

We do not got more insight on the choice of the local weight functions $G_{n,.}$ and f_n because they have small influences in practice, either in our functional framework as in standard multivariate bandwidth selection problems (see Vieu 1991).

4 Asymptotic optimality of the bandwidth selectors

Let us first introduce a set of assumptions under which our asymptotic results can be obtained. These assumptions will be widely commented later along the Remark 1. Then, our main results will be presented in Sect. 4.2.

4.1 Assumptions

$$\text{The weight function } G \text{ is bounded and with support on } \mathcal{C}. \tag{4}$$

$$\begin{aligned}&\text{The weight functions } G_{n,t} \text{ are bounded and with compact supports in } \mathcal{B}\left(t, h^{\gamma}\right),\\ &\text{where } 0 < \gamma < 1 \text{ and we have denoted } \mathcal{B}\left(t, h^{\gamma}\right) = \left\{t' \in \mathcal{H}; d\left(t', t\right) < h^{\gamma}\right\}.\end{aligned} \tag{5}$$

$$\begin{aligned}&\mathcal{C} \text{ is some given compact subset of } \mathcal{H} \text{ having nonempty interior, such that}\\ &\mathcal{C} \subset \cup_{l=1}^{l_n} \mathcal{B}(z_l, r_n), \text{ where } l_n = r_n^{-\zeta} \text{ with } \zeta > 0, \text{ and } r_n \to 0 \text{ as } n \to \infty.\end{aligned} \tag{6}$$

$$\begin{aligned}&\text{The set } H_n\text{, where the bandwidths are selected, verifies that}\\ &\mathrm{card}(H_n) = O\left(n^{\tau_0}\right), \text{ where } \tau_0 > 0.\end{aligned} \tag{7}$$

$$\begin{aligned}&\text{The kernel function } K \text{ has support } [0, 1] \text{ and is Lipschitz continuous on } [0, \infty),\\ &\text{and there exists some } \theta \text{ such that for any } u \in [0, 1], \ -K'(u) > \theta > 0.\end{aligned} \tag{8}$$

$$\begin{aligned}&\text{The probability distribution of the infinite-dimensional process } T \text{ is assumed to}\\ &\text{satisfy, for any } h \in H_n, \ P(T \in \mathcal{B}(t, h)) = \mathcal{K}_t \phi(h) + o(\phi(h)), \text{ for any } t \in \mathcal{C},\end{aligned} \tag{9}$$

where

$$\sup_{t\in\mathcal{C}} \mathcal{K}_t < \infty, \int_0^1 \phi(hs)\,ds > \alpha_0\phi(h) \text{ with } \alpha_0 > 0, \phi(h) = O\left(n^{-\tau}\right) \text{ with } \tau > 0,$$

and

$$\lim_{h\to 0} \frac{\phi(hs)}{\phi(s)} > 0 \text{ for any } s \in [0, 1].$$

Let us introduce the following notation:

$$g_j(t) = E\left(X_j \mid T = t\right), \eta_j = X_j - E\left(X_j \mid T\right) \text{ and } \boldsymbol{\eta}_i = \left(\eta_{i1}, \ldots, \eta_{ip}\right)^T,$$

where $\{\eta_{ij}\}_i$ is an i.i.d. sample according to the distribution of η_j. Observe that the expressions of our estimators (2) and (3) contain estimators of $g_1, \ldots, g_p$. Thus, in addition to the usual smoothness conditions on m, we need similar ones on g_j. More precisely, we assume that all the operators to be estimated are smooth, in the sense that for some $C < \infty$ and some $\alpha > 0$ we have that:

$$\text{For any } (u, v) \in \mathcal{C}\times\mathcal{C} \text{ and } f \in \left\{m, g_1, \ldots, g_p\right\},\ |f(u) - f(v)| \le Cd(u, v)^\alpha. \tag{10}$$

Furthermore, we need the following assumptions:

$$\text{For any } k \in N^* \ (N^* \text{ denotes the set of natural numbers without the zero}), \text{ there exists } C_k > 0 \text{ such that } E(|\varepsilon|^k \mid T = t) < C_k, \text{ for any } t \in \mathcal{C}. \tag{11}$$

$$\text{There exists } \sigma > 0 \text{ such that } E(\varepsilon^2 \mid T = t) = \sigma(t) \ge \sigma, \text{ for any } t \in \mathcal{C}, \text{ where } \sigma(t) \text{ is continuous.} \tag{12}$$

$$\text{There exists } q \ge 3 \text{ such that } E\left(|\eta_1|^q\right) + \cdots + E\left(|\eta_p|^q\right) < \infty. \tag{13}$$

$$\mathbf{B} = E\left(\boldsymbol{\eta}_1\boldsymbol{\eta}_1^T\right) \text{ is a positive definite matrix,} \tag{14}$$

and

$$\boldsymbol{\eta}_i \text{ is independent of } \varepsilon_i \ (i = 1, \ldots, n). \tag{15}$$

Remark 1 The conditions above related to the nonparametric and/or functional settings, as well as those concerning the moments of the random errors, are usual conditions (see, for instance, Rachdi and Vieu 2007; Benhenni et al. 2007; Ferraty and Vieu 2008), while assumptions (13)–(15) are directly related to the semi-parametric modelling and are quite unrestrictive (see Aneiros-Pérez and Vieu 2006). Note that (14) is insuring the identifiability of the model, excluding situations when both variables X and T are linked in a non stochastic way. This identifiability condition is very usual in standard multivariate situations but avoids for considering interesting problems in functional data analysis when both X and T are associated to the same functional explanatory variable (further works in this direction are in progress). Note also that (6) could be overpassed by introducing sophisticated entropy conditions on the set $\mathcal{C}$, but this is not done here in order to keep clearer our main issues on bandwidth selection problems (details can be seen in Ferraty and Vieu (2009)). As told before, the exact form of the weight functions G and $G_{n,.}$ is of low importance, the main point being the natural restriction that $\gamma < 1$ which is obviously necessary to make sure that the local weight does not exclude more data than the bandwidth h itself does. Maybe, it should finally be emphasized the hypotheses which are related to the small ball probability functions ϕ. Indeed, these conditions are giving strong evidence for the necessity of using a very general abstract semi-metric modelling for the space $\mathcal{H}$. Looking at the wide variety of probabilistic results existing about asymptotic expression of small ball probability functions for various gaussian processes and/or for various norms (see Bogachev (1998), Li and Shao (2001), for a few results in this field), one could see that such a function ϕ will decrease to zero much faster than any polynomial rate. A first naive point of view would be to say that our set of assumptions is excluding all these standard processes because we need polynomial decreasing rate. However, this is true if one is dealing with standard metric/normed spaces while the using of general topology based on semi-metric makes possible to reach polynomial rate for any kind of infinitely dimensional processes (see Ferraty and Vieu (2006) for examples of such topology insuring polynomial decreasing as well as for their interest in applied studies). At the end, the negative naive point of view should be completely reconsidered and semi-metric modelling has to be seen as powerful statistical tool and not only as topological cosmetic.

4.2 Asymptotic optimality

We now give our asymptotic results. Theorem 1 studies the asymptotic optimality of the global bandwidth selector $\widehat{h}$, while Theorem 2 deals with the asymptotic optimality of the local bandwidth selector $\widehat{h}_t$. Both theorems will be proven in the Appendix.

Theorem 1 *Under Assumptions* (4) *and* (6)–(15)*, if in addition* $nh^{4\alpha} \to 0$ *and* $n^{4\tau_0} \sup_{h\in H_n} \phi(h) \log\log n \to 0$ *as* $n \to \infty$*, and* $\phi(h) \geq n^{(2/q)+b-1}/(\log n)^2$ *for* n *large enough and some constant* $b > 0$ *satisfying* $(2/q) + b > 1/2$*, then*

$$\frac{\mathrm{ASE}\left(\widehat{h}\right)}{\inf_{h\in H_n} \mathrm{ASE}\,(h)} \to 1 \text{ a.s. as } n \to \infty.$$

Theorem 2 *If in the enunciate of Theorem 1 we use Assumption (5) instead of Assumption (4), then*

$$\frac{\mathrm{ASE}_t\left(\widehat{h}_t\right)}{\inf_{h\in H_n}\mathrm{ASE}_t\left(h\right)} \to 1 \text{ a.s. as } n \to \infty.$$

5 A simulation study

A simulation study was designed to observe how our bandwidth selectors perform when the sample size is finite, and to compare the global and local procedures. Two settings were considered. On the one hand, we dealt with smooth curves and moderate variability in the errors. On the other hand, the case of rough curves and high variability in the errors was studied.

Because our bandwidth selectors are constructed to estimate the regression function r (equivalently, to estimate the nonparametric component m), all the measures below concern this estimation setting.

5.1 Smooth curves

5.1.1 The simulated models

The SFPLR model

$$Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + m(T_i) + \varepsilon_i \quad (i = 1, \ldots, n)$$

was considered, where X_{ij} and ε_i were i.i.d. according to a $N(0, 1)$ and a $N(0, \sigma_\varepsilon)$, respectively (we have denoted $\sigma_\varepsilon = 0.1\,(\max_T m(T) - \min_T m(T))$). The functional data were $T_i(z) = a_i(z - 0.5)^2 + b_i$ $(z \in [0, 1])$ with a_i and b_i being i.i.d. according to a $U(0, 1)$ and a $U(-0.5, 0.5)$, respectively. The curves T_i's were discretized on the same grid generated from 50 equispaced points in [0, 1]. The unknown vector of parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ was $\beta = (-1, 3)^T$, while two models were considered for the unknown function m; say Model M1, where

$$m(T_i) = \exp\left(-8f(T_i)\right) - \exp\left(-12f(T_i)\right)$$

with

$$f(T_i) = sign\left(T_i'(1) - T_i'(0)\right)\sqrt{3\int_0^1 \left(T_i'(z)\right)^2 dz},$$

and Model M2, where

$$m(T_i) = dnorm\left(0.5\int_0^1 T_i''(z)dz; 0.5, 0.05\right)$$

with $dnorm\ (x; \mu, \sigma)$ being the normal density with mean μ and standard deviation σ taken at point x. Note that the expressions above give $m(T_i) = \exp(-8a_i) - \exp(-12a_i)$ and $m(T_i) = dnorm\ (a_i; 0.5, 0.05)$ for Models M1 and M2, respectively (see Staniswalis (1989) for a simulation study considering these kinds of regression functions in a pure nonparametric finite dimensional setting; that is, eliminating the linear component in the SFPLR model and using the scalar argument a_i instead of the curve T_i).

For each model we simulated various sample sizes $n = 50, 100, 200$, and for each of them the same experiment was replicated $M = 100$ times.

5.1.2 Choosing the parameters of the estimates

The Epanechnikov kernel was used in the estimates. The proximity between the curves T_i was taken into account to construct the set H_n where the bandwidths were selected (more precisely, the set H_n is the interval defined from the quantiles 0.05 and 0.5 of the empirical distribution of $d(T_i, T_j), i \neq j$). In addition, the weight function $G\ (T_i) = 1$ was considered for the global procedure, while for the local we took a weight function in the class

$$G_{n,t}\ (T_i) = \begin{cases} 1, & \text{if } d\ (T_i, t) < k_n \\ 0, & \text{otherwise} \end{cases}$$

where the parameter k_n was selected by means of cross-validation. Once again (see Remark 1), let us recall that the form of these weights has low practical impact.

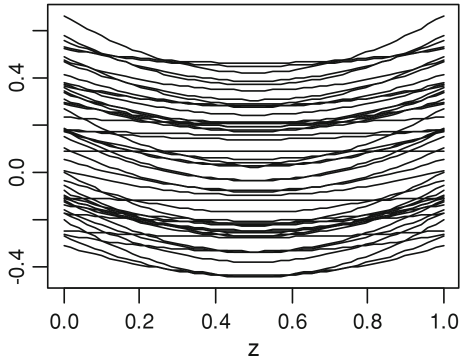
The smoothness of the curves T_i (see Fig. 1) lead us to consider semi-metrics based on the L_2 norm of some derivative of the curves (see Ferraty and Vieu 2006, Chap. 3). Several orders (0, 1, 2 and 3) for the derivative were considered, and we chose the order which gave the best predictions on a sample test. Thus, we used the semi-metric

$$d\left(T, T^*\right) = \left(\int_0^1 \left(T'(z) - T^{*\prime}(z)\right)^2 dz\right)^{1/2} = \left(\frac{1}{12}\int_0^1 \left(T''(z) - T^{*\prime\prime}(z)\right)^2 dz\right)^{1/2}$$

for both Models M1 and M2 (the computation of this semi-metric was based on B-spline approximation of each curve). Note that the last equality above is a consequence of the kind of curves considered. Also, the selected order is in accordance with the features of the function m corresponding to each model considered.

In this study we aim to look at the behavior of the two bandwidth selection rules defined in the sections above: the global data-driven bandwidth h minimizing the

Fig. 1 Sample of 50 curves T_i's (smooth curves)



criterion CV, and the local bandwidth h_t minimizing at each new curve t the local criterion CV_t.

5.1.3 Cross-validated and optimal theoretical bandwidths

In a first attempt we studied how these data-driven bandwidths approximate the best (theoretical) bandwidths. For this, we consider the ratios:

$$R^G_{n,k} = \frac{\text{ASE}\left(\widehat{h}_n\right)}{\inf_{h\in H_n} \text{ASE}\,(h)}, \quad k = 1, \ldots, M,$$

and

$$R^L_{n,k,T^0_j} = \frac{\text{ASE}_{T^0_j}\left(\widehat{h}_{T^0_j,n}\right)}{\inf_{h\in H_n} \text{ASE}_{T^0_j}\,(h)} \quad k = 1, \ldots, M, \;\; j = 1, \ldots, 100,$$

where $\left\{\left(Y^0_j, X^0_{j1}, X^0_{j2}, T^0_j\right)\right\}_{j=1}^{100}$ was a test sample. Boxplots of $\left\{R^G_{n,k}\right\}_k$ and $\left\{R^L_{n,k,T^0_j}\right\}_k$ are reported in Figs. 2 and 3, respectively (for the local case, these boxplots are shown only for two values of j). These figures show the presence of some outliers, but both their number and size decrease as the sample size n increases. Furthermore, a considerable decrease to 1 in the ratios as n increases is suggested.

The information given by Figs. 2 and 3 is summarized in Tables 1 and 2, respectively, which show the mean and the standard deviation of $\left\{R^G_{n,k}\right\}_k$ and $\left\{R^L_{n,k,T^0_j}\right\}_k$, respectively (for the local case, these descriptive measures are shown only for four values of j).

In conclusion (from Figs. 2 and 3 and Tables 1 and 2), the good behavior of the bandwidth selectors is clear. The global (respectively, local) cross-validated bandwidth produces a global (respectively, local) estimation error which is close to the minimal one. The good point is that the asymptotic behavior stated in Theorems 1 and 2 turns

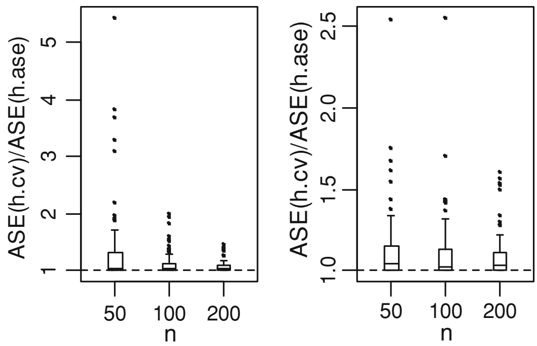


Fig. 2 Boxplots of the ratios of the ASE's when global bandwidths are considered. *Left side*: Model M1. *Right side*: Model M2

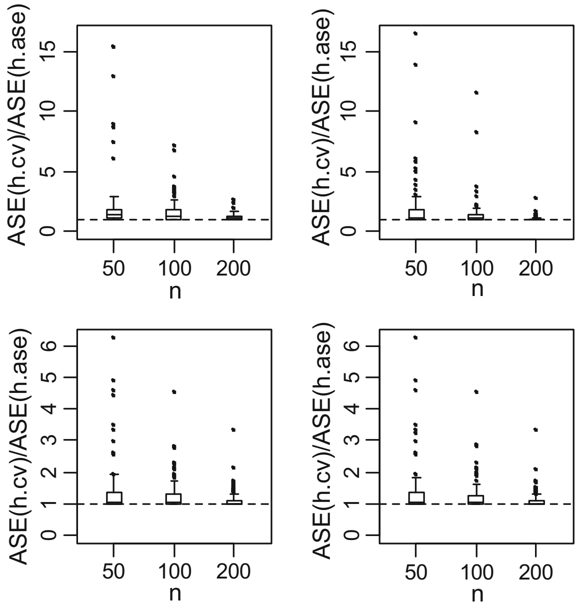


Fig. 3 Boxplots of the ratios of the ASE's when local bandwidths are used. *Upper panel*: Model M1. *Lower panel*: Model M2 (two curves are considered for each model)

Table 1 Mean and standard deviation (in parenthesis) corresponding to the ratios of the ASE's when global bandwidths are considered

n	Model	
	M_1	M_2
50	1.3049 (0.6682)	1.1151 (0.2071)
100	1.1177 (0.1943)	1.1037 (0.1971)
200	1.0676 (0.0965)	1.0858 (0.1296)

to be apparent already for moderate sample sizes, at least for the specific models considered in this study. In addition, note that the speed of convergence to 1 in the ratio corresponding to the global selector is fastest than that of the local selector.

Table 2 Mean and standard deviation (in parenthesis) corresponding to the ratios of the ASE's when local bandwidths are considered (these descriptive measures are shown for four curves)

n	Model	
	M_1	M_2
50	2.1731 (2.5739)	1.4179 (0.8931)
	1.9330 (2.2892)	1.4056 (0.8594)
	2.1686 (2.5754)	1.4496 (0.9325)
	2.3785 (2.9818)	1.3864 (0.7807)
100	1.6616 (1.0289)	1.2579 (0.4990)
	1.4631 (1.3175)	1.2403 (0.4928)
	1.6623 (1.0286)	1.2469 (0.5119)
	1.6634 (1.0261)	1.2795 (0.5915)
200	1.2310 (0.3207)	1.1363 (0.3024)
	1.0722 (0.2078)	1.1310 (0.2978)
	1.2338 (0.3322)	1.1420 (0.2937)
	1.2314 (0.3377)	1.1488 (0.2939)

5.1.4 Predictive behavior of cross-validated estimates

In a second attempt, it is worth looking at the main point of interest for people in practice: what is the impact of this good behavior on the prediction? So, we completed the study as followings. For each sample size and model considered, the last replication was taken for predicting the values $\left\{Y_j^0\right\}_{j=1}^{100}$. In this way, global and local bandwidths were used. The accuracy of the predictions was measured by using the Mean Squared Errors of Prediction (MSEP)

$$\mathrm{MSEP}_n^{\mathrm{G}} = \frac{\sum_{j=1}^{100}\left(Y_j^0 - \widehat{r}_{\widehat{h}_n}\left(X_{j1}^0, X_{j2}^0, T_j^0\right)\right)^2}{100}$$

and

$$\mathrm{MSEP}_n^{\mathrm{L}} = \frac{\sum_{j=1}^{100}\left(Y_j^0 - \widehat{r}_{\widehat{h}_{T_j^0,n}}\left(X_{j1}^0, X_{j2}^0, T_j^0\right)\right)^2}{100},$$

corresponding to the global and local selectors, respectively. The values of these measures, and their ratios, are given in Table 3.

In conclusion, the results reported in Table 3 indicate that, in general, for the models considered in this simulation study, the predictions obtained as well with local bandwidths as with a global one are rather good. In this situation, local procedure gives slightly best results but in counter part they need heavier computational time. Note also that to be able to capture local features of the data a rather high sample size

Table 3 Accuracy measure MSEP using global (G) and local (L) bandwidths, together with their ratios L/G

n	Model			
	M_1 [a]		M_2	
50	4.6424^G	1.0492	1.3507	0.8929
	4.8706^L		1.2060	
100	2.9158	0.9848	0.9552	0.9280
	2.8713		0.8864	
200	2.6043	0.9060	0.7862	0.8944
	2.3596		0.7032	

[a] The MSEPs (but not their ratios) corresponding to Model M1 have been multiplied by 10^4

is needed, since the interest of local selection procedure starts to be significant for $n = 200$.

5.2 Rough curves

5.2.1 The simulated model

To give more evidence of the interest of our methodology, we proceeded a second kind of simulated experiments, based on data having much higher variability than those used previously. This is done from one side by generating rough curves in the following way

$$T_i(z) = a_i \sin(4(b_i - z)) + b_i + \vartheta_{i,z} \quad (z \in [0, 1]),$$

with a_i, b_i and $\vartheta_{i,z}$ being r.v. i.i.d. according to a $N(4, 3)$, $N(0, 3)$ and $N(0, 0.5)$ distributions, respectively. These curves are generated in a nonparametric way (because of the random term ϑ) and, as depicted in Fig. 4, they present much higher variability than the previous ones presented in Fig. 1.

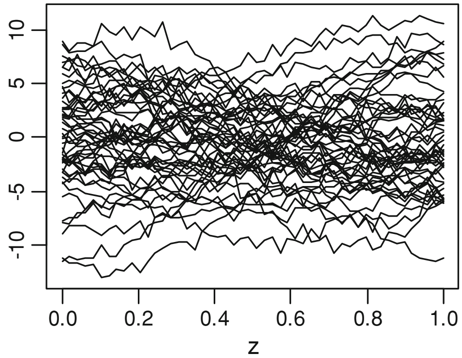


Fig. 4 Sample of 50 curves T_i's (rough curves)

From another hand we have also increased the variability by choosing higher variance for the innovation error ε,

$$\sigma_\varepsilon = 0.5 \left(\max_T m(T) - \min_T m(T) \right),$$

and finally, we have considered the following regression operator

$$m(T_i) = \int_0^1 \frac{dz}{1 + |T_i(z)|}.$$

5.2.2 The parameters of the estimates

When dealing with rough curves such as those introduced herein it does not make sense to measure the proximity by means of a semi-metric based on the derivatives of the curves. We used rather a new family of semi-metrics, based on the first components of the functional principal component analysis of the data, and the number of principal components was selected by using cross-validation techniques. Further motivations as well as full details on the using of such functional PCA based semi-metrics can be found in Ferraty and Vieu (2006).

All other parts of the experiment (choice of the kernel K, choice of the local and global weight functions $G_{n,.}$ and G, choice of the sample sizes n and construction of the learning and testing samples) remain the same as they were before for smooth curves.

5.2.3 The results

The next Figs. 5 and 6 show boxplots of the ratios of the ASE errors when global and local cross-validation procedures are used, respectively (that is, boxplots of the ratios $\{R^G_{n,k}\}_k$ and $\{R^L_{n,k,T^0_j}\}_k$, respectively).

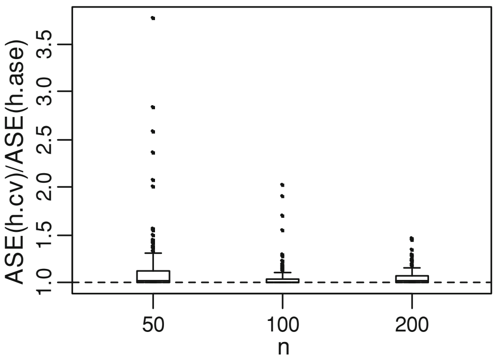


Fig. 5 Boxplots of the ratios of the ASE's when global bandwidths are considered

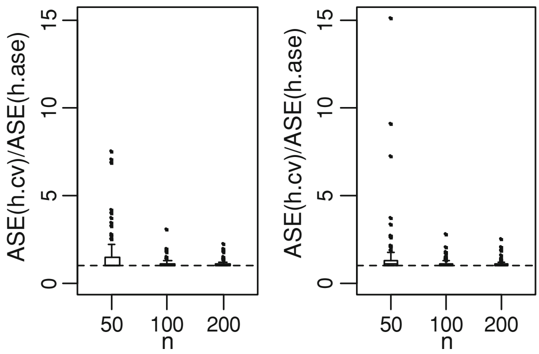


Fig. 6 Boxplots of the ratios of the ASE's when local bandwidths are used (two curves are considered)

Table 4 Mean and standard deviation (in parenthesis) corresponding to the ratios of the ASE's when global or local bandwidths are considered (for the local case those measures are shown for four curves)

n	Global	Local
50	1.1694 (0.4156)	1.5549 (1.1985)
		1.5358 (1.7676)
		1.6331 (2.9564)
		1.6497 (3.0202)
100	1.0633 (0.1648)	1.1274 (0.2637)
		1.1197 (0.2438)
		1.0873 (0.1880)
		1.0897 (0.2348)
200	1.0571 (0.0907)	1.0979 (0.2001)
		1.1042 (0.2417)
		1.0707 (0.1288)
		1.0566 (0.0972)

Table 5 Accuracy measure MSEP using global (G) and local (L) bandwidths, together with their ratios L/G

n					
50		100		200	
0.1122^{G}	1.0916	0.1164	1.1831	0.1057	1.1275
0.1225^{L}		0.1377		0.1192	

As observed before in Figs. 2 and 3 for smooth curves, one can see that both global and local cross-validation procedures work well. To emphasize more on that, we present in Table 4 the mean and the standard deviations of these ratios.

Once again, as observed before in Tables 1 and 2, one gets empirical evidence of the asymptotic optimality of both bandwidth selection methods that was stated in the theoretical part of this paper. Even for rather moderated sample size ($n = 100$), the ratio between cross-validated ASE and optimal theoretical ASE is closed to 1. The impact in terms of prediction is shown through Table 5.

As a matter of conclusion, one could say that both selection procedures (local and global) have good predictive power. In this specific example there is a slight advantage for the global procedure but the local one gives very similar results.

6 Concluding remarks

In this paper, we constructed two bandwidth selectors for the estimator of the regression function in a SFPLR model. These selectors are based on global and local cross-validation ideas, respectively. The asymptotic optimality of the selectors was obtained, and their finite sample size behavior, as well as a comparison between them, were illustrated by means of a simulation study.

The various simulation studies showed the good behavior of the bandwidth selectors proposed. From one side, the ratios of the ASE's (evaluated in both the cross-validation and the optimal bandwidths) go to 1 when moderate sample sizes increase, confirming the fact that the cross-validated bandwidths are closed to the optimal theoretical ones. More importantly in practice, it has also been observed that both (local and global) procedures have good predictive powers.

Appendix

Notation

The proofs of Theorems 1 and 2 are strongly linked with the corresponding results obtained by Rachdi and Vieu (2007) and Benhenni et al. (2007), respectively, who work on a pure nonparametric functional regression model. We introduce the following notation:

$$Y_j^* = m\left(T_j\right) + \varepsilon_j = Y_j - \sum_{k=1}^{p} X_{jk}\beta_k,$$

$$\widehat{m}_h^*(t) = \sum_{j=1}^{n} w_{n,h}(t, T_j)Y_j^*, \ \widehat{m}_h^{*i}(t) = \sum_{j\neq i}^{n} w_{n,h}^i(t, T_j)Y_j^*,$$

$$\mathrm{CV}^*(h) = n^{-1}\sum_{i=1}^{n}\left(Y_i^* - \widehat{m}_h^{*i}(T_i)\right)^2 G(T_i),$$

$$\mathrm{ASE}^*(h) = n^{-1}\sum_{i=1}^{n}\left(\widehat{m}_h^*(T_i) - m(T_i)\right)^2 G(T_i),$$

$$\mathrm{MISE}^*(h) = \int E\left(\left(\widehat{m}_h^*(t) - m(t)\right)^2\right) G(t)\, dP_T(t),$$

where P_T is the probability distribution measure of the functional variable T,

$$\widehat{\mathbf{m}}_h^* = \left(\widehat{m}_h^*(T_1), \ldots, \widehat{m}_h^*(T_n)\right)^T, \ \widehat{\mathbf{m}}_h^{*(-)} = \left(\widehat{m}_h^{*1}(T_1), \ldots, \widehat{m}_h^{*n}(T_n)\right)^T$$

and

$$\mathbf{Y}^* = \left(Y_1^*, \ldots, Y_n^*\right)^T.$$

Furthermore, we will denote

$$\mathbf{m} = (m(T_1), \ldots, m(T_n))^T , \ \widehat{\mathbf{m}}_h = (\widehat{m}_h(T_1), \ldots, \widehat{m}_h(T_n))^T ,$$
$$\widehat{\mathbf{m}}_h^{(-)} = \left(\widehat{m}_h^1(T_1), \ldots, \widehat{m}_h^n(T_n)\right)^T , \ \widehat{\mathbf{g}}_h^{(-)} = \left(\widehat{g}_{j,h}^i(T_i)\right)_{1\le i\le n;\ 1\le j\le p} ,$$

where $\widehat{g}_{j,h}^i(t) = \sum_{k\neq i}^n w_{n,h}^i(t, T_k) X_{jk}$, and

$$\mathbf{X}_h^{(-)} = \mathbf{X} - \widehat{\mathbf{g}}_h^{(-)} \quad \text{and} \quad \mathbf{W}_h^{(-)} = \left((1-\delta_{ij}) w_{n,h}^i \left(T_i, T_j\right)\right)_{1\le i\le n;\ 1\le j\le n} ,$$

where δ_{ij} is the Dirac delta function (note that $\mathbf{X}_h^{(-)} = \left(\mathbf{I} - \mathbf{W}_h^{(-)}\right)\mathbf{X}$). Finally, $\mathbf{G}$ denotes the $(n \times n)$ diagonal matrix whose i-th diagonal element is $G(T_i)$.

Auxiliary results

Before we begin with the proofs of our theorems, we will state some auxiliary results which play a main role in these proofs.

Lemma 1 (Aneiros-Pérez and Vieu 2006) *Under Assumptions* (6)*,* (8)–(10) *and* (13) (*m not included in* (10))*, if in addition $\left\{\left(X_{i1}, \ldots, X_{ip}, T_i\right)\right\}_{i=1}^n$ are independent and distributed as $\left(X_1, \ldots, X_p, T\right)$, and $h \to 0$ and $\log n / (n\phi(h)) \to 0$ as $n \to \infty$, we have that*

$$n^{-1}\widetilde{\mathbf{X}}_h^T \widetilde{\mathbf{X}}_h \longrightarrow \mathbf{B} \ a.s.$$

Theorem 3 (Aneiros-Pérez and Vieu 2006) *Under Assumptions* (6)*,* (8)–(15) *and* (13)*, if in addition $nh^{4\alpha} \to 0$ as $n \to \infty$ and $\phi(h) \ge n^{(2/q)+b-1} / (\log n)^2$ for n large enough and some constant $b > 0$ satisfying $(2/q) + b > 1/2$, then*

$$\lim\sup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} \left|\widehat{\boldsymbol{\beta}}_{hj} - \beta_j\right| = \left(\sigma_\varepsilon^2 b^{jj}\right)^{1/2} \ a.s.,$$

where $\sigma_\varepsilon^2 = Var(\varepsilon)$ and $b^{jj} = \left(\mathbf{B}^{-1}\right)_{jj}$.

Lemma 2 *Under Assumptions* (6), (8)–(10) *and* (13) (*m not included in* (10))*, if in addition $\left\{\left(X_{i1}, \ldots, X_{ip}, T_i\right)\right\}_{i=1}^n$ are independent and distributed as $\left(X_1, \ldots, X_p, T\right)$, and $h \to 0$ and $\log n / (n\phi(h)) \to 0$ as $n \to \infty$, we have that*

$$n^{-1}\mathbf{X}_h^{(-)T}\mathbf{X}_h^{(-)} \longrightarrow \mathbf{B} \ a.s.$$

Proof of Lemma 2 This proof can be obtained in a similar way as that of Lemma 1, and therefore it is omitted.

Proof of Theorem 1

Rachdi and Vieu (2007) obtained that

$$\frac{\text{ASE}^*\left(\widehat{h}^*\right)}{\inf_{h\in H_n}\text{ASE}^*(h)} \to 1 \text{ a.s. as } n \to \infty,$$

where we have denoted

$$\widehat{h}^* = \arg\min_{h\in H_n} \text{CV}^*(h)$$

(at this moment, we should indicate that the kernel function used in Rachdi and Vieu (2007) does not verify our Assumption (8) but, as they noted, the result above remains under (8). We have used this assumption because we need the uniform convergence of the nonparametric estimators; see Ferraty and Vieu 2008). Therefore, we have that the proof of Theorem 1 is complete if we prove both

$$\sup_{h\in H_n}\left|\frac{\text{ASE}(h)-\text{ASE}^*(h)}{\text{ASE}^*(h)}\right| \to 0 \text{ a.s. as } n\to\infty \tag{16}$$

and

$$\sup_{h\in H_n}\left|\frac{\text{CV}(h)-\text{CV}^*(h)}{\text{CV}^*(h)}\right| \to 0 \text{ a.s. as } n\to\infty. \tag{17}$$

For this, we will use the decompositions

$$\begin{aligned}\text{ASE}(h) = \text{ASE}^*(h) &+ n^{-1}\left(\widehat{\boldsymbol{\beta}}_h-\boldsymbol{\beta}\right)^T\widetilde{\mathbf{X}}_h^T\mathbf{G}\widetilde{\mathbf{X}}_h\left(\widehat{\boldsymbol{\beta}}_h-\boldsymbol{\beta}\right)\\ &+2n^{-1}\left(\widehat{\mathbf{m}}_h^*-\mathbf{m}\right)^T\mathbf{G}\widetilde{\mathbf{X}}_h\left(\widehat{\boldsymbol{\beta}}_h-\boldsymbol{\beta}\right)\end{aligned} \tag{18}$$

and

$$\begin{aligned}\text{CV}(h) = \text{CV}^*(h) &+ n^{-1}\left(\widehat{\boldsymbol{\beta}}_h-\boldsymbol{\beta}\right)^T\mathbf{X}_h^{(-)T}\mathbf{G}\mathbf{X}_h^{(-)}\left(\widehat{\boldsymbol{\beta}}_h-\boldsymbol{\beta}\right)\\ &-2n^{-1}\left(\mathbf{Y}^*-\widehat{\mathbf{m}}_h^{*(-)}\right)^T\mathbf{G}\mathbf{X}_h^{(-)}\left(\widehat{\boldsymbol{\beta}}_h-\boldsymbol{\beta}\right),\end{aligned} \tag{19}$$

respectively, where we have considered the facts that

$$\widehat{\mathbf{m}}_h = \widehat{\mathbf{m}}_h^* - \mathbf{W}_h\mathbf{X}\left(\widehat{\boldsymbol{\beta}}_h-\boldsymbol{\beta}\right)$$

and

$$\widehat{\mathbf{m}}_h^{(-)} = \widehat{\mathbf{m}}_h^{*(-)} - \mathbf{W}_h^{(-)}\mathbf{X}\left(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}\right),$$

respectively.

Proof of (16) From (18) together with the Cauchy-Schwarz inequality we have that

$$\begin{aligned}\left|\mathrm{ASE}\,(h) - \mathrm{ASE}^*\,(h)\right| &\leq n^{-1}\left|\left(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}\right)^T \widetilde{\mathbf{X}}_h^T \mathbf{G}\widetilde{\mathbf{X}}_h \left(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}\right)\right| \\ &\quad + 2n^{-1}\left|\left(\widehat{\mathbf{m}}_h^* - \mathbf{m}\right)^T \mathbf{G}\widetilde{\mathbf{X}}_h \left(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}\right)\right| \\ &\leq n^{-1}\left|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}\right|^2 \left\|\widetilde{\mathbf{X}}_h\right\|^2 \|\mathbf{G}\| \\ &\quad + 2n^{-1}\left|\left(\widehat{\mathbf{m}}_h^* - \mathbf{m}\right)^T \mathbf{G}^{1/2}\right| \left\|\mathbf{G}^{1/2}\right\| \left\|\widetilde{\mathbf{X}}_h\right\| \left|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}\right|.\end{aligned}$$

Thus, taking into account the facts that $\left|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}\right| = O_{\text{a.s.}}\left(\left(n^{-1}\log\log n\right)^{1/2}\right)$ (see Theorem 3), $\left\|\widetilde{\mathbf{X}}_h\right\| = O_{\text{a.s.}}\left(n^{1/2}\right)$ (see Lemma 1) and $\|\mathbf{G}\| = O(1)$ (see Assumption (4)), together with both the asymptotic equivalence between the quadratic measures $\mathrm{ASE}^*\,(h)$ and $\mathrm{MISE}^*\,(h)$ and the fact that $\mathrm{MISE}^*\,(h) \geq C\,(n\phi\,(h))^{-1}$ (Rachdi and Vieu 2007), we obtain

$$\sup_{h\in H_n}\left|\frac{\mathrm{ASE}\,(h) - \mathrm{ASE}^*\,(h)}{\mathrm{ASE}^*\,(h)}\right| = O\left(\max\left\{a_{n,\phi}, a_{n,\phi}^{1/2}\right\}\right) \text{ a.s.,} \tag{20}$$

where we have denoted $a_{n,\phi} = card^4\,(H_n)\sup_{h\in H_n}\phi\,(h)\log\log n$. Finally, Assumption (7) together with the fact that $n^{4\tau_0}\sup_{h\in H_n}\phi\,(h)\log\log n \to 0$ as $n\to\infty$ give (16).

Proof of (17) Rachdi and Vieu (2007) obtained that

$$\sup_{h\in H_n}\left|\frac{\mathrm{CV}^*\,(h) - A_n - \widetilde{\mathrm{ASE}}^*\,(h)}{\widetilde{\mathrm{ASE}}^*\,(h)}\right| \to 0 \text{ a.s. as } n \to \infty,$$

where $A_n = n^{-1}\sum_{i=1}^n \varepsilon_i^2 G\,(T_i)$ and $\widetilde{\mathrm{ASE}}^*\,(h) = n^{-1}\sum_{i=1}^n \left(\widehat{m}_h^{*i}\,(T_i) - m(T_i)\right)^2 G\,(T_i)$, this last function being a quadratic measure asymptotically equivalent to $\mathrm{ASE}^*\,(h)$ (and therefore to $\mathrm{MISE}^*\,(h)$). Thus, we have that

$$\mathrm{CV}^*\,(h) \text{ is asymptotically equivalent to } \mathrm{MISE}^*\,(h) + A_n. \tag{21}$$

Now, a similar reasoning as that followed to obtain (20) (using (19) and Lemma 2 instead of (18) and Lemma 1, respectively, and taking into account both (21) and the fact that $A_n \geq 0$) gives

$$\sup_{h\in H_n}\left|\frac{\mathrm{CV}\,(h) - \mathrm{CV}^*\,(h)}{\mathrm{CV}^*\,(h)}\right| = O\left(\max\left\{a_{n,\phi}, a_{n,\phi}^{1/2}\right\}\right) \text{ a.s.}$$

Finally, as in the case of (16), we have that (17) holds.

Proof of Theorem 2

This proof is analogous to that of Theorem 1, but using the results of Benhenni et al. (2007) instead of those of Rachdi and Vieu (2007).

Acknowledgements Philippe Vieu wishes to thank all the participants of the work group STAPH on Functional Statistics at the University Paul Sabatier of Toulouse for stimulating and helpful comments, and more specifically Frédéric Ferraty for his continuous fruitful collaboration on nonparametric FDA. The activities of this group are available on http://www.lsp.ups-tlse.fr/staph. We are also grateful to two anonymous referees for their constructive comments and suggestions, which helped to improve the quality of the paper. The research of Germán Aneiros-Pérez was supported by Xunta de Galicia Grant PGIDIT07PXIB105259PR.

References

Aneiros-Pérez G, Quintela-del-Río A (2001) Modified cross-validation in semiparametric regression models with dependent errors. Commun Stat Theory Methods 30:289–307

Aneiros-Pérez G, Vieu P (2006) Semi-functional partial linear regression. Stat Probab Lett 76:1102–1110

Benhenni K, Ferraty F, Rachdi M, Vieu P (2007) Local smoothing regression with functional data. Comput Stat 22:353–369

Bogachev VI (1998) Gaussian measures. Mathematical surveys and monographs 62. American Mathematical Society

Delsol L (2009) Advances on asymptotic normality in nonparametric functional time series analysis. Statistics 43:13–33

Engle R, Granger C, Rice J, Weiss A (1986) Nonparametric estimates of the relation between weather and electricity sales. J Am Stat Assoc 81:310–320

Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer, New York

Ferraty F, Vieu P (2008) Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. J Nonparametr Stat 20:187–189 (Erratum of J Nonparametr Stat 16:111–125, 2004)

Ferraty F, Vieu P (2009) Kernel regression estimation for functional data. In: Ferraty F, Romain Y (eds) Handbook on statistics for functional and operatorial statistics. Oxford University Press, Oxford (to appear)

Härdle W, Marron JS (1985) Optimal bandwidth choice in nonparametric regression function estimation. Ann Stat 13:1465–1481

Härdle W, Liang H, Gao J (2000) Partially linear models. Springer-Verlag, New York

Li WV, Shao QM (2001) Gaussian processes: inequalities, small ball probabilities and applications. In: Rao CR, Shanbhag D (eds) Stochastic processes: theory and methods. Handbook of Statistics, vol 19. North-Holland, Amsterdam, pp 533–597

Linton O (1995) Second order approximation in the partially linear regression model. Econometrica 63:1079–1112

Rachdi M, Vieu P (2007) Nonparametric regression for functional data: automatic smoothing parameter selection. J Stat Plan Inference 137:2784–2801

Ramsay J, Silverman B (1997) Functional data analysis. Springer, New York

Ramsay J, Silverman B (2002) Applied functional data analysis. Springer, New York

Ramsay J, Silverman B (2005) Applied functional data analysis, 2nd edn. Springer, New York

Robinson P (1988) Root-n-consistent semiparametric regression. Econometrica 56:931–954

Speckman P (1988) Kernel smoothing in partial linear models. J R Stat Soc Ser B 50:413–436

Staniswalis JG (1989) Local bandwidth selection for kernel estimates. J Am Stat Assoc 84:284–288

Vieu P (1991) Nonparametric regression: optimal local bandwidth choice. J R Stat Soc Ser B 53:453–464