UNIVERSIDADE DE
SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE
ESTATÍSTICA E INVESTIGACIÓN OPERATIVA

SiZer Map for Evaluating a Bootstrap Local Bandwidth Selector in Nonparametric Additive Models

M. D. Martínez-Miranda, R. Raya-Miranda,
W. González-Manteiga and A. González-Carmona

Report 05-01

Reports in Statistics and Operations Research

SiZer Map for Evaluating a Bootstrap Local Bandwidth Selector in Nonparametric Additive Models

M.D. Martínez-Miranda, R. Raya-Miranda,
W. González-Manteiga and A. González-Carmona*

Abstract

We propose a bootstrap local bandwidth selector for additive models. The selector is derived from a bootstrap approximation of the conditional mean squared error, based on a wild bootstrap resampling scheme over the estimated residuals. The selector is computed exactly (without involving Monte Carlo approximations) and in practice can be evaluated for many additive estimation methods, including backfitting (bivariate), marginal integration and mixed methods. We study the consistency of the bootstrap approximation and also carry out an empirical simulation study to explore the performance of the proposed selector in comparison with others, considering estimation with backfitting. The graphical tool SiZer Map enables us to make meaningful comparisons between local and global selectors.

Key Words: Wild Bootstrap, Smoothing Parameter, Backfitting, Marginal Integration.

1. INTRODUCTION

Additive models in nonparametric regression were first suggested by Friedman and Stuetzle (1981) and were popularised by Hastie and Tibshirani (1990). This additive modelling provides an elegant solution to the so called *curse of dimensionality*, which arises with all the multivariate extensions of smoothing techniques (kernel regression, local linear regression, etc.). In fact, the problem

*M.D. Martínez-Miranda is Associate Professor (email: mmiranda@ugr.es); R. Raya-Miranda is Assistant Professor (email: rraya@ugr.es); A. González-Carmona is Professor (email: andresgc@ugr.es), Department of Statistics and O.R., Faculty of Sciences, University of Granada, Campus Fuentenueva, 18071, Granada, Spain; and W. González-Manteiga is Professor (email: wenceslao@usc.es), Department of Statistics and O.R., Faculty of Mathematics, University of Santiago de Compostela, Campus South, 15706, Santiago de Compostela, Spain.
This work was partially supported by MCyT Grants PB98-0182-C02-02 and BFM2002-03213 (European FEDER support included) and XUGA Grant PGIDT00PXI20704PN.

can be avoided by separating the effect of each covariate X_d over the response variable, Y, as additive models do by considering the following regression model,

$$Y = \alpha + \sum_{d=1}^{D} m_d(X_d) \tag{1}$$

where α is a constant term and $\{m_d,\, d = 1, \ldots, D\}$ denotes a set of unknown univariate functions (called component functions). Moreover, studies by Stone (1985, 1986) reveal that such additive models can be estimated by achieving the univariate convergence rate, this being independent of the fixed number of covariates involved.

The backfitting algorithm for estimating additive models (Buja, Hastie and Tibshirani, 1989) is widely accepted because of its straightforward implementation and intuitive definition. However the theoretical properties of backfitting estimates are poorly understood because of their iterative origin. These theoretical difficulties have motivated the development of other estimation methods in recent years. An alternative estimation method, which has received considerable interest, is that of the marginal integration, introduced by Linton and Nielsen (1995). Far from competing against the above powerful techniques, the latter seeks to achieve the best solution by combining their good features. Thus, Linton (1997) and Kim, Linton and Hengartner (1999) proposed such mixed methods, using starting estimations provided by the marginal integration, within the backfitting algorithm. The estimators thus derived are more efficient than those obtained by applying pure backfitting. Sperlich, Linton and Härdle (1999) carried out an interesting comparative study between the two techniques, showing that, in general, neither can be definitively considered superior to the other; their behaviour depends on the model considered and even on the estimation point inside the estimation interval. Backfitting performs better at boundary points and also when the correlation between the covariates is high. Many of the theoretical problems of backfitting were solved by Opsomer and Ruppert (1997), who considered local polynomial smoothers in a bivariate context. The well-known properties of these smoothers make it possible to derive the asymptotic properties of backfitting solutions and, moreover, to provide explicit expressions for them. A multivariate extension of the asymptotic properties of backfitting estimators was derived by Opsomer (2000), but this author did not provide explicit expressions for them such as were given for the bidimensional situation. Instead, a recursive expression for the additive smoother was proposed, which was useful for these asymptotic purposes.

The advantages of using local polynomial smoothing are not restricted to backfitting algorithms; Severance-Lossin and Sperlich (1997) proved the consistency and derived the asymptotic properties of the estimators by marginal integration when these smoothers are involved. Furthermore, the estimators are robust for choosing bandwidths, and also reduce bias and thus the mean squared error.

The choice of smoothing parameter (or parameters) becomes a crucial technical problem in nonparametric and semiparametric regression, because of its direct repercussion on the behaviour of smoothers. For additive models (1) estimated by a backfitting procedure, it features bandwidth selectors based on crossvalidation or generalized crossvalidation (Kauermann and Opsomer, 2004), and also a simple selector called *rule of thumb* proposed by Linton and Nielsen (1995). Besides these automatic selectors, others proposed have been based on plug-in methodology; see, for example, Opsomer and Ruppert (1998) and Severance-Lossin and Sperlich (1997).

Crossvalidation selectors are attractive due to their simplicity and intuitive definition, but they suffer from two main disadvantages (as described by Opsomer and Ruppert, 1998) that have been widely investigated and accepted not only in additive modelling but in the general context of nonparametric curve estimation. From a theoretical perspective, crossvalidation provides estimators with a convergence rate that is limited to $O_P(n^{-1/10})$, for the unidimensional case; moreover, they have a high sample variability (Härdle, Hall and Marron, 1988). Besides, in practice and from a computational point of view, crossvalidation is too expensive. Many authors, including Gu and Wahba (1988), have developed more efficient variants, but even with these improvements, the procedure requires a long time because of the considerable number of additive fittings.

By contrast, plug-in methodologies for selecting smoothing parameters require us to obtain theoretical expressions (normally asymptotic) of the bias and the variance of regression estimators. Such expressions are not always available and achieving them can be difficult because of the very definition of estimators, as happens when they derive from iterative processes such as a backfitting algorithm. In addition, these asymptotic expressions are restricted to severe theoretical conditions on the model, its design and of course they must be valid for large sample sizes.

On the other hand, both automatic selectors and plug-in selectors choose smoothing parameters by minimizing error criteria such as MASE (mean averaged squared error), MISE (mean integrated squared error) and ASE (averaged squared error). The global nature of these measures leads to smoothing parameter values being constant over the whole estimation interval.

Local bandwidth selectors for nonparametric smoothers provide notable improvements in the estimation of surfaces by achieving a major adaptation to the subjacent features of data (Fan and Gijbels, 1995, introduced such a local selector for univariate local polynomial smoothers based on plug-in ideas). We have no information about their use for nonparametric estimating additive models. For semiparametric additive models only, the paper by Opsomer and Ruppert (1999) defined a bandwidth selector that extended the local one proposed by Ruppert (1997). Nevertheless, any simple local selectors, like the local version of the crossvalidation selector proposed by Vieu (1991), for a unidimensional nonparametric regression context, can be easily adapted to the settings assumed in the present paper. Without assuming additivity, González-Manteiga, Martínez-Miranda and Pérez-González (2004) adapted local crossvalidation for selecting bandwidths in order to estimate bidimensional regression surfaces by local linear smoothers in their simulations.

Another outstanding tool for selecting bandwidths, in nonparametric curve estimation, is the bootstrap methodology. For additive modelling (or generalized additive models), many papers have used bootstrap for constructing confidence intervals for regression surfaces (Kauermann and Opsomer, 2003; Kim, Linton and Hengartner, 1999), and for solving tests (Sperlich, Tjϕstheim and Yang, 2002, formulated tests for detecting interactions between covariates by using wild bootstrap; also Yang, Sperlich and Härdle, 2003, used wild bootstrap for testing parametric models). In these papers, wild bootstrap is applied to resample the regression residuals when the marginal integration procedure is considered, under both additive models and generalized additive models with a known link. In the same context, Dalelane (1999) and Härdle, Huet, Mammen and Sperlich (2001) (for generalized additive models) derive the consistency of the wild bootstrap procedure.

In the present paper, we propose an estimation of the conditional mean squared error, based on wild bootstrap, for the problem of estimating a D-variate additive model. The bootstrap approximation arises from this general multidimensional context, while the available explicit expressions for backfitting estimators involving local linear smoothers (as proposed by Opsomer and Ruppert, 1997), for the bivariate case, allow us to compute the bootstrap approximations exactly, in practice. It also happens when marginal integration and also other mixed methods are considered. The key for this lies in the linear form of the estimators over the responses ($\widehat{m}(\mathbf{x}) = \sum_{i=1}^{n} w_i(\mathbf{x})Y_i$), which lead to the following simple expressions for the conditional bias and variance:

$$\begin{aligned} \mathrm{B}(\widehat{m}(\mathbf{x})) &= \sum_{i=1}^{n} w_i(\mathbf{x})m(\mathbf{X}_i) - m(\mathbf{x}), \\ \mathrm{V}(\widehat{m}(\mathbf{x})) &= \sum_{i=1}^{n} w_i^2(\mathbf{x})\sigma^2(\mathbf{X}_i). \end{aligned} \tag{2}$$

This property allows us to compute in a straightforward way the exact expressions of the proposed bootstrap approximations, without needing Monte Carlo approximations, which would involve considerable computational cost.

Such a bootstrap approximation of the conditional mean squared error enables us to define a local bandwidth selector which can be extended immediately to other estimation techniques, like marginal integration and methods which combine marginal integration with backfitting. Note that, in such cases, the consistency of wild bootstrap was derived by Dalelane (1999) and Kim et al. (1999), respectively. Moreover, since the derivation of exact explicit expressions of bootstrap approximations is based on the linearity of regression estimators, in these cases the local bootstrap bandwidth selector obtained could be calculated again, without involving Monte Carlo techniques.

This suggestion for bootstrap estimation follows the same methodology as was proposed by González-Manteiga, Martínez-Miranda and Pérez-González (2004) for the multivariate local linear estimator. These authors obtained a selector which performed very well in estimating regression surfaces. In the additive modelling we assume, the local bootstrap selector again performs very well, as shown in the simulation experiments described.

The remainder of this paper is structured as follows: the next section describes the notation and preliminary definitions. Section 3 contains the resampling mechanism for constructing the bootstrap approximation of the conditional mean squared error and the definition of the local bootstrap bandwidth selector. By assuming different additive estimation methods, we calculate exact expressions of bootstrap bias and variance, and also demonstrate the consistency of bootstrap. Section 4 describes a simulation study for exploring the finite sample properties of the bootstrap approximations. The exploratory tool, called SiZer Map, introduced by Chaudhuri and Marron (1999) and adapted to additive models by Raya-Miranda, Martínez-Miranda and González-Carmona (2002), is used in this section to evaluate the behaviour of the proposed bandwidth selector. SiZer allows us to make meaningful comparisons with other well-known selectors. In Section 5, some conclusions are drawn, and finally the proofs are briefly summarized.

2. NOTATION AND PRELIMINAR DEFINITIONS

Let us assume that

$$Y_i = \alpha + \sum_{d=1}^{D} m_d(X_{di}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3}$$

where $\{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)\}$, with $\mathbf{X}_i = (X_{1i}, \ldots, X_{Di})^T$, is a set of independent and identically distributed random variables as a $\mathbb{R}^{D+1}$-valued variable $(\mathbf{X}, Y)$. The residuals, ε_i, are independent and identically distributed with mean 0 and different variances, $\sigma^2(\mathbf{X}_i)$. To ensure the identifiability of the component functions, $m_d(\cdot)$, we include the intercept, α, and assume $\mathrm{E}[m_d(X_{di})] = 0$ $(d = 1, \ldots, D)$.

Let $f(\mathbf{x})$ denote the density of vector $\mathbf{X}$ and $f_d(x_d)$ $(d = 1, \ldots, D)$, the marginal densities.

2.1. Estimation by Backfitting

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^T$. We write the vectors of the additive functions at the observation points as $\mathbf{m}_d = (m_d(X_{d1}), \ldots, m_d(X_{dn}))^T$, $d = 1, \ldots, D$. Under the additive regression model (3) represents the $n \times n$ smoother matrix with respect to the dth covariate vector as $\mathbf{S}_d$. The component functions, $\mathbf{m}_d$, can be estimated nonparametrically at the observation points by solving the following system of normal equations

$$\begin{bmatrix} \mathbf{I} & \mathbf{S}_1 & \cdots & \mathbf{S}_1 \\ \mathbf{S}_2 & \mathbf{I} & \cdots & \mathbf{S}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_D & \mathbf{S}_D & \cdots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_D \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_D \end{bmatrix} \mathbf{Y} \tag{4}$$

The backfitting algorithm (Buja, Hastie and Tibshirani, 1989) provides an iterative solution of (4). When a bivariate additive model is assumed and the smoother matrices $\mathbf{S}_d$ are based on local polynomial regression estimators, Opsomer and Ruppert (1997) derived explicit expressions of the backfitting estimators. We show these expressions in the following by introducing this additional notation:

For any bidimensional point of estimation, $\mathbf{x} = (x_1, x_2)^T$, let s_{d,x_d}^T represent the equivalent kernels for the local polynomial regression at x_d with a degree p_d $(d = 1, 2)$. These equivalent kernels can be written as

$$s_{d,x_d}^T = e_1^T \left(\mathbf{X}_{x_d}^T \mathbf{W}_{x_d} \mathbf{X}_{x_d}\right)^{-1} \mathbf{X}_{x_d}^T \mathbf{W}_{x_d}$$

where the $p_d + 1$-dimensional vector $e_1^T = (1, 0, \ldots, 0)$, the matrix

$$\mathbf{W}_{x_d} = \operatorname{diag}\left\{\frac{1}{h_d} K\left(\frac{X_{d1} - x_d}{h_d}\right), \ldots, \frac{1}{h_d} K\left(\frac{X_{dn} - x_d}{h_d}\right)\right\},$$

for a kernel function K and the Vector of smoothing parameters $\mathbf{h} = (h_1, h_2)^T$, and

$$\mathbf{X}_{x_d} = \begin{pmatrix} 1 & (X_{d1} - x_d) & \cdots & (X_{d1} - x_d)^{p_d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_{dn} - x_d) & \cdots & (X_{dn} - x_d)^{p_d} \end{pmatrix}. \tag{5}$$

Let the smoother matrices be $\mathbf{S}_d$ with the rows being the equivalent kernels at the observations, X_{di} $(i = 1, \ldots, n;\ d = 1, 2)$, respectively. We define the vector of fitted values at the observation points as

$$\widehat{\mathbf{m}} = \widehat{\alpha} + \widehat{\mathbf{m}}_1 + \widehat{\mathbf{m}}_2,$$

with the constant α being estimated by $\overline{\mathbf{Y}}$, and $\widehat{\mathbf{m}}_d$ $(d = 1, 2)$ being the solutions to the set of estimating equations

$$\begin{pmatrix} \mathbf{I} & \mathbf{S}_1^* \\ \mathbf{S}_2^* & \mathbf{I} \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{m}}_1 \\ \widehat{\mathbf{m}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{S}_1^* \\ \mathbf{S}_2^* \end{pmatrix} \mathbf{Y},$$

where $\mathbf{S}_d^* = \left(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n\right) \mathbf{S}_d$ $(d = 1, 2)$ and $\mathbf{1}$ is the n-valued vector $(1, \ldots, 1)^T$.

In such a bivariate context, the backfitting algorithm gives the solutions with the following explicit expressions:

$$\begin{aligned} \widehat{\mathbf{m}}_1 &= \left\{\mathbf{I} - \left(\mathbf{I} - \mathbf{S}_1^*\mathbf{S}_2^*\right)^{-1} \left(\mathbf{I} - \mathbf{S}_1^*\right)\right\} \mathbf{Y} \equiv \mathbf{W}_1 \mathbf{Y} \\ \widehat{\mathbf{m}}_2 &= \left\{\mathbf{I} - \left(\mathbf{I} - \mathbf{S}_2^*\mathbf{S}_1^*\right)^{-1} \left(\mathbf{I} - \mathbf{S}_2^*\right)\right\} \mathbf{Y} \equiv \mathbf{W}_2 \mathbf{Y} \end{aligned} \tag{6}$$

provided the inverses also exist.

For $\widehat{\mathbf{m}}$ we have

$$\widehat{\mathbf{m}} = \left\{\mathbf{1}\mathbf{1}^T/n + 2\mathbf{I} - \left(\mathbf{I} - \mathbf{S}_1^*\mathbf{S}_2^*\right)^{-1} \left(\mathbf{I} - \mathbf{S}_1^*\right) - \left(\mathbf{I} - \mathbf{S}_2^*\mathbf{S}_1^*\right)^{-1} \left(\mathbf{I} - \mathbf{S}_2^*\right)\right\} \mathbf{Y} \equiv \mathbf{W}\mathbf{Y}.$$

For simplicity, we assume $\overline{\mathbf{y}} = 0 = \widehat{\alpha}$ in the following.

The estimation (6) at any estimation point, $\mathbf{x} = (x_1, x_2)^T$, can be obtained simply by applying another step of the backfitting algorithm, which yields the following expressions:

$$\widehat{m}(\mathbf{x}) = \widehat{m}_1(x_1) + \widehat{m}_2(x_2) \doteq \widehat{m}_{\mathbf{h}}^{BA}(\mathbf{x}) \tag{7}$$

where

$$\widehat{m}_1(x_1) = s_{1,x_1}^T(\mathbf{Y} - \widehat{\mathbf{m}}_{\mathbf{2}}) = s_{1,x_1}^T(\mathbf{Y} - \mathbf{W}_2\mathbf{Y}) \tag{8}$$

and similarly

$$\widehat{m}_2(x_2) = s_{2,x_2}^T(\mathbf{Y} - \widehat{\mathbf{m}}_{\mathbf{1}}) = s_{2,x_2}^T(\mathbf{Y} - \mathbf{W}_1\mathbf{Y}), \tag{9}$$

with $\mathbf{W_1}$ and $\mathbf{W_2}$ having been defined in (6). Thus, the estimator is linear over the responses with weights being explicitly written by

$$w_i^{BA}(\mathbf{x}, \mathbf{h}) = s_{1,x_1}^T \left(\mathbf{I} - \mathbf{S}_2^*\mathbf{S}_1^*\right)^{-1} \left(\mathbf{I} - \mathbf{S}_2^*\right) + s_{2,x_2}^T \left(\mathbf{I} - \mathbf{S}_1^*\mathbf{S}_2^*\right)^{-1} \left(\mathbf{I} - \mathbf{S}_1^*\right), \tag{10}$$

for each estimation point, $\mathbf{x}$.

2.2. Marginal Integration and mixed methods

The marginal integration method described by Linton and Nielsen (1995) and Hengartner (1996) among others, estimates each component $m_d(\cdot)$ by integrating a pilot multivariate smoother of $m(\cdot)$ involving a $D-1$-dimensional probability measure. The so-called empirical marginal integration estimator is defined by means of a D-dimensional Nadaraya-Watson estimator, as follows:

Let us consider the partition $\mathbf{X}_i \equiv (X_{di}, \mathbf{X}_{-d,i})$ where $\mathbf{X}_{-d,i}$ is the $D-1$-dimensional vector defined by removing X_{di} from $\mathbf{X}_i$. By analogy, $\mathbf{x} \equiv (x_d, \mathbf{x}_{-d})$. The multivariate Nadaraya-Watson's estimator is given by

$$\widetilde{m}(\mathbf{x}) = \frac{1}{nh^D} \sum_{i=1}^{n} K_D\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right) \frac{Y_i}{\widehat{f}(\mathbf{x})}$$

where $K_D(t_1, \ldots, t_D) = \prod_{d=1}^{D} K(t_d)$ with $K(\cdot)$ a scalar kernel, h a scalar bandwidth, and $\widehat{f}(\cdot)$ an estimator of the density $f(\cdot)$.

Thus, the empirical marginal integration estimator of the d-th component, $m_d(\cdot)$, is computed by

$$\widehat{\gamma}_d(x_d) = n^{-1} \sum_{j=1}^{n} \widetilde{m}(x_d, \mathbf{X}_{-d,j}) \doteq \sum_{i=1}^{n} w_i^{dIM}(x_d, h) Y_i,$$

with weights

$$w_i^{dIM}(x_d, h) = n^{-2} h^{-D} \widehat{f}(x)^{-1} \sum_{j=1}^{n} K\left(\frac{x_d - X_{dj}}{h}\right) K_{D-1}\left(\frac{\mathbf{X}_{-d,i} - \mathbf{X}_{-d,j}}{h}\right). \tag{11}$$

The additive reconstruction is then

$$\begin{aligned}\widehat{m}^{IM}(\mathbf{x}) &= \sum_{d=1}^{D}\widehat{\gamma}_d(x_d) = \sum_{i=1}^{n}\left\{\sum_{d=1}^{D} w_i^{dIM}(x_d,h)\right\}Y_i \\ &\doteq \sum_{i=1}^{n} w_i^{IM}(\mathbf{x},h)Y_i \doteq \widehat{m}_h^{IM}(\mathbf{x}).\end{aligned} \tag{12}$$

Kim et al. (1999) proposed an efficient oracle estimator which was defined by inserting the empirical marginal integration estimator (12) into a backfitting algorithm but taking one step only. To define this estimator, first the adjusted responses are constructed by

$$Y_{di}^{2-step} = Y_i - \sum_{j\neq d}^{D}\widehat{\gamma}_j(X_{ji},h_0) \tag{13}$$

with h_0 being a pilot scalar bandwidth. Now we apply the univariate local polynomial smoother of the pairs $\{(X_{di},Y_{di}^{2-step}), i=1,\dots n\}$ to estimate the d-th component, m_d, which can be written as

$$\widehat{m}_d^{2-step}(x_d) = \sum_{i=1}^{n} w_i^{LP}(x_d,h_d)Y_{di}^{2-step}, \tag{14}$$

where h_d is a scalar bandwidth, and the weights w_i^{LP} are those associated to the univariate local polynomial smoother (of degree p_d), i.e.,

$$\begin{aligned} w_i^{LP}(x_d,h_d) = \quad & e_1^T\left(\mathbf{X}_{x_d}^T\widetilde{\mathbf{W}}_{x_d}\mathbf{X}_{x_d}\right)^{-1}\times \\ & \times\left[1,(X_{di}-x_d),\dots,(X_{di}-x_d)^p\right]^T K_{h_d}(X_{di}-x_d)\mathbf{1}_{X_{di}\in S^{on}}\end{aligned}$$

with a boundary correction that also affects the elements of $\widetilde{\mathbf{W}}_{x_d}$, by defining sets $S^{on}=\left\{\mathbf{x}\in\mathbb{R}^D : \underline{b}_d+h_0\le x_d\le\overline{b}_d-h_0,\ d=1,\dots,D\right\}$ (where $\underline{b}_d$ and $\overline{b}_d$ are, respectively, the lower and upper bounds of the support of X_{di}, which for simplicity is assumed to be rectangular).

By substituting (13) in (14) we obtain

$$\widehat{m}_d^{2-step}(x_d) = \sum_{i=1}^{n} w_i^{dOR}(x_d,h_d,h_0)Y_i$$

with

$$w_i^{dOR}(x_d,h_d,h_0) = w_i^{LP}(x_d,h_d) - \sum_{l=1}^{n} w_l^{LP}(x_d,h_d)\sum_{j=1(j\neq d)}^{D} w_i^{jIM}(X_{jl},h_0). \tag{15}$$

Thus the additive estimator given by Kim et al. (1999) can be written in a linear way over the responses as follows:

$$\begin{aligned}\widehat{m}^{2-step}(\mathbf{x}) = & \sum_{i=1}^{n}\left\{\sum_{d=1}^{D} w_i^{dOR}(x_d, h_d, h_0)\right\} Y_i \\ & \doteq \sum_{i=1}^{n} w_i^{OR}(\mathbf{x}, \mathbf{h}, h_0) Y_i \doteq \widehat{m}^{2-step}_{\mathbf{h},h_0}(\mathbf{x}),\end{aligned} \tag{16}$$

with a D-dimensional vector of bandwidths $\mathbf{h} = (h_1, \ldots, h_D)^T$ and also a scalar bandwidth h_0.

2.3. The local error criterion

To evaluate the additive estimation procedures, we consider the local error criterion, named the conditional Mean Squared Error (MSE) and which is given by

$$\mathrm{MSE}\,(\mathbf{x}; \mathbf{h}) = \mathrm{E}_{Y|\mathbf{X}}\left[\{\widehat{m}\,(\mathbf{x}) - m\,(\mathbf{x})\}^2\right], \tag{17}$$

where $\mathbf{h}$ is the vector of smoothing parameters involved in the estimation $\widehat{m}(\mathbf{x})$.

Under the above defined error criterion, we can now define the optimum theoretical local vector of smoothing parameters, denoted by $\mathbf{h}_{opt}(\mathbf{x})$, as follows:

$$\mathbf{h}_{opt}(\mathbf{x}) = \arg\min_{\mathbf{h}} \mathrm{MSE}(\mathbf{x}; \mathbf{h}), \tag{18}$$

for each estimation point, $\mathbf{x}$.

3. WILD BOOTSTRAP FOR ADDITIVE MODELS

3.1. Bootstrapping the conditional Mean Squared Error

The bootstrap methodology known as wild bootstrap has been shown to perform well in estimating a nonparametric regression model in a situation of heterocedasticity as is assumed in this paper (3). Wild bootstrap was introduced by Wu (1986) in the context of linear regression models and has recently been used by González-Manteiga et al. (2004) in the context of multivariate local linear regression, in order to introduce a local multivariate-bandwidth selector. The resampling mechanism considered here follows the same guidelines as those that inspired the methodology proposed by these authors, but assumes additivity in the multivariate regression model.

Let us consider the following resampling scheme:

i) Estimate the residuals by $\widehat{\varepsilon}_i = Y_i - \widehat{m}_0(\mathbf{X}_i)$, where $\widehat{m}_0$ represents a pilot additive estimator of the regression function.

ii) Generate the bootstrap residuals, ε_i^*, $i = 1, ..., n$, verifying $\mathrm{E}^*\left[\varepsilon_i^*\right] = 0$, $\mathrm{E}^*\left[\varepsilon_i^{*2}\right] = \widehat{\varepsilon}_i^2$ and $\mathrm{E}^*\left[\varepsilon_i^{*3}\right] = \widehat{\varepsilon}_i^3$.

iii) Draw the bootstrap sample as $\left\{\left(\mathbf{X}_i^T, Y_i^*\right)\right\}$, defining the bootstrap responses as $Y_i^* = \widehat{m}_0(\mathbf{X}_i) + \varepsilon_i^*$, $i = 1, ..., n$.

iv) Calculate the bootstrap additive estimator, $\widehat{m}^*(\cdot)$, based on the bootstrap sample generated in the previous step.

Under such specifications, the bootstrap approximation of the conditional mean squared error on each estimation point $\mathbf{x}$ is defined by

$$\mathrm{MSE}^*(\mathbf{x}) = \mathrm{E}^*\left[\left(\widehat{m}^*(\mathbf{x}) - \widehat{m}_0(\mathbf{x})\right)^2\right]. \tag{19}$$

Remark 1 *The resampling mechanism is defined for any additive estimator (such as those defined in section 2) and considers within the pilot estimator,* $\widehat{m}_0(\cdot)$, *a pilot vector of smoothing parameters,* $\mathbf{g}$, *a different one from that involved in the fourth step, denoted by* $\mathbf{h}$. *Thus the bootstrap approximation (19) is denoted by* $MSE^*(\mathbf{x}; \mathbf{h}, \mathbf{g})$ *in the following.*

Remark 2 *A similar bootstrap method was defined by Kim et al. (1999) for their efficient oracle estimator (16) but involving in the first step a pilot estimator like* $\widehat{m}_0(\mathbf{x}) \equiv \widehat{m}^{2-step}_{\mathbf{h},h_0}(\mathbf{x})$ *involving a vector bandwidth* $\mathbf{h}$ *different from the one considered in the estimator introduced in the third step (used to build the bootstrap responses) which are denoted by* $\mathbf{g}$. *Moreover, the same pilot scalar bandwidth,* h_0 *(involved in the empirical marginal integration estimator) is applied for all the estimators used. With such a mechanism, these authors proved the consistency of bootstrap for defining bootstrap confidence intervals for the regression function. The authors also concluded that a pilot bandwidth* $\mathbf{g}$ *of larger order than* $n^{-1/(2q+1)}$ *(with* q *a measure of the smoothness of* m*) is needed. These oversmoothing conditions for the pilot estimator will be necessary again for consistency when backfitting estimators are involved (see Theorem 1).*

The above proposed bootstrap approximation (19) allows us to define a bandwidth selector that is local, i.e., that is a function of the estimation point, $\mathbf{x}$, because of the locality of the measure of error considered.

Let us define the bootstrap local bandwidth selector as the function

$$\mathbf{h}^*_{boot}(\mathbf{x}) = \arg\min_{\mathbf{h}} \mathrm{MSE}^*(\mathbf{x}; \mathbf{h}, \mathbf{g}) \tag{20}$$

on a given estimation point, $\mathbf{x}$.

3.2. Exact expressions for bootstrap approximations

The bootstrap methodology and the approximations derived are defined in the general D-dimensional additive regression context and for any estimation method, and they are accessible in practice by following such a resampling method and by considering the Monte Carlo approximations as usual. Nevertheless, many of these estimation methods allow us to compute exact expressions that are available (in practice) for the bootstrap approximation (19), without considering those computationally expensive procedures.

In the following, we derive these exact expressions for three methods defined in section 2 and also study the consistency of the bootstrap mechanism.

3.2.1. Backfitting with $D = 2$

The exact expressions available (in practice) for the backfitting regression estimators, introduced by Opsomer and Ruppert (1997) for the bivariate case (when local linear smoothers are involved), given at (6), allowed us to prove the consistency of the bootstrap approximation (19), and also to obtain an expression that can be evaluated (in practice), by computing, exactly, the expectation over the resampling distribution. Thus it is possible to achieve the practical implementation of the proposed local bootstrap bandwidth selector (20) without involving Monte Carlo approximations.

Assume the bidimensional regression context and the backfitting regression estimates considered by these authors, as described in Section 2. In the following we give a result which establishes the consistency of the bootstrap approximation (19) in such a context. For this purpose, we first introduce some additional notation and formulate a set of assumptions.

Associated to the kernel function, K, the j-th moment is defined by $\mu_j(K) = \int u^j K(u)du$, and the quantity $R(K) = \int K(u)^2 du$.

Assumptions:

A1. The kernel K is bounded and continuous, it has compact support and its first derivative has a finite number of sign changes over its support. Also, $\mu_j(K) = 0$ for all odd j, $\mu_2(K) \neq 0$.

A2. The densities f, f_1, f_2 are bounded and continuous, have compact support and their first derivatives have a finite number of sign changes over their supports. Also, $f_d(x_d) > 0$ $(d = 1, 2)$ for all $\mathbf{x} = (x_1, x_2)^t \in \text{supp}(f)$ and

$$\sup \left\{ \left| \frac{f(x_1, x_2)}{f_1(x_1) f_2(x_2)} - 1 \right|, (x_1, x_2)^t \in \text{supp}(f) \right\} < 1. \tag{21}$$

A3. As $n \to \infty$, $h_d \to 0$ and $nh_d/\text{log}n \to \infty$ $(d = 1, 2)$. Analogous for g_1 and g_2.

A4. For each point $\mathbf{x} \in \text{supp}(f)$, the function σ^2 is strictly positive and continuously differentiable, the densities f, f_d $(d = 1, 2)$ and the regression function m are twice continuously differentiable.

A5. The function $\nu_4(\mathbf{x}) = \text{E}\left[Y^4 \mid \mathbf{X} = \mathbf{x}\right]$ is uniformly bounded on $\mathbf{x}$.

A6. $h_d g_d^{-1} \to 0$ $(d = 1, 2)$ as $n \to \infty$.

The condition (21) in assumption A2 allows us to obtain an asymptotic representation of the weights of the backfitting regression estimator (7). The asymptotic representation is based on lemmas 3.1 and 3.2 in Opsomer and Ruppert (1997) and is shown in the following preliminary result:

Lemma 1. Under assumptions A1 to A3, the backfitting estimator involving local linear smoothers (7) admits an asymptotic approximation given by

$$\widehat{m}_{\mathbf{h}}^{BA}(\mathbf{x}) \approx n^{-1} f_1^{-1}(x_1) \sum_{i=1}^{n} K_{h_1}(X_{1i} - x_1) Y_i + n^{-1} f_2^{-1}(x_1) \sum_{i=1}^{n} K_{h_2}(X_{2i} - x_2) Y_i$$
$$+ o_P(n^{-1} h_1^{-1}) + o_P(n^{-1} h_2^{-1}) + o_P(n^{-2} h_1^{-1} h_2^{-1}), \tag{22}$$

on each estimation point $\mathbf{x}$.

Now let us establish the consistency of the proposed bootstrap approximation.

Theorem 1. Under assumptions A1–A6, the bootstrap approximation MSE* is consistent in the interior of supp(f), i.e.,

$$\mathrm{MSE}^*(\mathbf{x}; \mathbf{h}, \mathbf{g}) - \mathrm{MSE}(\mathbf{x}; \mathbf{h}) \to 0$$

in probability.

Remark 3 *The expression (22) implies that, from an asymptotic perspective, the effects of each covariate can be considered separately, and so the estimator is defined as a sum of univariate local linear regression estimators (note that it is necessary to add the estimation of constant α to these weights if noncentering additive models are considered). This feature allows us to demonstrate the consistency in a way similar to that used by González-Manteiga et al. (2004).*

Remark 4 *Condition A6 means that the bootstrap approximation must involve a vector of pilot bandwidths oversmoothing the regression function. This characteristic appears again for bootstrap approximation in another context, in fact, when unidimensional Nadaraya-Watson regression estimators are considered, Härdle and Marron (1991) impose an order of $n^{-1/9}$ for the pilot bandwidth. In the multivariate context, González-Manteiga et al. (2004) generalize this requirement to the order $n^{-1/(8+D)}$, considering local linear regression.*

Remark 5 *(**Exact expression of bootstrap MSE**) The simple resampling distribution considered by wild bootstrap allows straightforward calculation of the expectation E^*, to derive exact expressions for the bootstrap mean squared error approximation. These exact expressions avoid the need to use Monte Carlo techniques that would involve very long computing times, especially considering point approximations as proposed here. Next, we derive these exact expressions.*

Consider a decomposition of the conditional mean squared error into a bias component given by

$$\mathrm{B}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x}) = \mathrm{E}^*\left[\widehat{m}_{\mathbf{h}}^{BA*}(\mathbf{x})\right] - \widehat{m}_{\mathbf{g}}^{BA}(\mathbf{x}),$$

called the bootstrap bias, and the above-mentioned bootstrap variance:

$$\mathrm{V}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x}) = \mathrm{Var}^*\left(\widehat{m}_{\mathbf{h}}^{BA*}(\mathbf{x})\right).$$

Using the linearity of the bias and the variance (2) the bootstrap bias can be written as

$$\mathrm{B}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x}) = \sum_{i=1}^{n} w_i^{BA}(\mathbf{x},\mathbf{h})\,\widehat{m}_{\mathbf{g}}^{BA}(\mathbf{X}_i) - \widehat{m}_{\mathbf{g}}^{BA}(\mathbf{x}), \tag{23}$$

and the bootstrap variance as

$$\mathrm{V}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x}) = \sum_{i=1}^{n} w_i^{BA}(\mathbf{x},\mathbf{h})^2 \left(Y_i - \widehat{m}_{\mathbf{g}}^{BA}(\mathbf{X}_i)\right)^2. \tag{24}$$

The following equivalent expressions can be obtained by using matrix notation, substituting the expressions for the weights (10):

$$\begin{aligned}
\mathrm{B}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x}) = \; & n^{-1}\Big(\mathbf{1} + s^T_{1,x_1,h_1}\left(\mathbf{I}-\mathbf{S}_2^*\mathbf{S}_1^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_2^*\right) + \\
& + s^T_{2,x_2,h_2}\left(\mathbf{I}-\mathbf{S}_1^*\mathbf{S}_2^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_1^*\right)\Big)\widehat{M}_{\mathbf{g}} - \\
& - n^{-1}\Big(\mathbf{1} + s^T_{1,x_1,g_1}\left(\mathbf{I}-\mathbf{S}_2^*\mathbf{S}_1^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_2^*\right) + \\
& + s^T_{2,x_2,g_2}\left(\mathbf{I}-\mathbf{S}_1^*\mathbf{S}_2^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_1^*\right)\Big)\mathbf{Y}
\end{aligned}$$

$$\begin{aligned}
\mathrm{V}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x}) = \; & n^{-2}\Big(\mathbf{1} + s^T_{1,x_1,h_1}\left(\mathbf{I}-\mathbf{S}_2^*\mathbf{S}_1^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_2^*\right) + \\
& + s^T_{2,x_2,h_2}\left(\mathbf{I}-\mathbf{S}_1^*\mathbf{S}_2^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_1^*\right)\Big)\Sigma_{\mathbf{g}} \times \\
& \times \Big(\mathbf{1} + s^T_{1,x_1,h_1}\left(\mathbf{I}-\mathbf{S}_2^*\mathbf{S}_1^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_2^*\right) + \\
& + s^T_{2,x_2,h_2}\left(\mathbf{I}-\mathbf{S}_1^*\mathbf{S}_2^*\right)^{-1}\left(\mathbf{I}-\mathbf{S}_1^*\right)\Big)
\end{aligned}$$

where

$$\widehat{M}_{\mathbf{g}} = \begin{pmatrix} \widehat{\alpha} + \widehat{m}_{1;g_1}(X_{11}) + \widehat{m}_{2;g_2}(X_{21}) \\ \vdots \\ \widehat{\alpha} + \widehat{m}_{1;g_1}(X_{1n}) + \widehat{m}_{2:g_2}(X_{2n}) \end{pmatrix} \quad \text{and} \quad \Sigma_{\mathbf{g}} = \mathrm{diag}\left(\widehat{\varepsilon}_i\right).$$

Remark 6 (**Bootstrap bandwidth selector**) *The above obtained exact expressions (23) and (24) can be used to calculate the proposed local bootstrap bandwidth selector (20). With these expressions, the selector is intuitive and simple to calculate in practice. Furthermore, the computational implementation of the selector in empirical studies is not very expensive in comparison with other bootstrap versions based on the Monte Carlo technique, and also in comparison with other local bandwidth selectors such as one based on crossvalidation that was introduced by Vieu (1991). More details about comparisons and computational times are given with the simulation experiments described below.*

3.2.2. Marginal integration and mixed methods

By arguments similar to those applied to the backfitting estimators above, the bootstrap approximation (19) can be evaluated exactly in practice, when the empirical marginal integration estimator (12) and the efficient oracle estimator (16) are used within the resampling mechanism described above. The expressions are similar to (23) and (24) for bootstrap bias and variance, respectively, but involving the corresponding new weights ((11) and (16)), i.e.,

$$\mathrm{B}_{h,g}^{*}\left(\mathbf{x}\right)=\sum_{i=1}^{n}w_{i}^{IM}\left(\mathbf{x},h\right)\widehat{m}_{g}^{IM}\left(\mathbf{X}_{i}\right)-\widehat{m}_{g}^{IM}\left(\mathbf{x}\right),$$

and

$$\mathrm{V}_{h,g}^{*}\left(\mathbf{x}\right)=\sum_{i=1}^{n}w_{i}^{IM}\left(\mathbf{x},h\right)^{2}\left(Y_{i}-\widehat{m}_{g}^{IM}\left(\mathbf{X}_{i}\right)\right)^{2},$$

for the empirical marginal integration, with scalar bandwidths h and g. For the efficient oracle estimator of Kim et al. (1999),

$$\mathrm{B}_{\mathbf{h},\mathbf{g},h_0}^{*}\left(\mathbf{x}\right)=\sum_{i=1}^{n}w_{i}^{OR}(\mathbf{x},\mathbf{h},h_0)\widehat{m}_{\mathbf{g},h_0}^{2-step}\left(\mathbf{X}_{i}\right)-\widehat{m}_{\mathbf{g},h_0}^{2-step}\left(\mathbf{x}\right),$$

$$\mathrm{V}_{\mathbf{h},\mathbf{g},h_0}^{*}\left(\mathbf{x}\right)=\sum_{i=1}^{n}w_{i}^{OR}(\mathbf{x},\mathbf{h},h_0)^{2}\left(Y_{i}-\widehat{m}_{\mathbf{g},h_0}^{2-step}\left(\mathbf{X}_{i}\right)\right)^{2},$$

with vector bandwidths $\mathbf{h},\mathbf{g}$ and a scalar bandwidth h_0. Kim et al. (1999) gave efficient guidelines for implementing the estimators and weights involved, thus enabling straightforward computation of the above proposed bootstrap selector, in practice, again for these additive estimation methods.

4. EMPIRICAL STUDY WITH BACKFITTING

Let us now investigate the performance of the proposed bootstrap approximation of the conditional mean squared error and the local bootstrap bandwidth selector based on it, in order to estimate additive models like (1) by backfitting. A notable element in this process is the graphical analysis made using SiZer Map, the graphical tool introduced by Chaudhuri and Marron (1999) for a unidimensional context. SiZer Map was extended and adapted to be applied to the estimation of nonparametric additive models, like those assumed in this paper, by Raya-Miranda et al. (2002).

Our objective is to evaluate the behaviour of certain bandwidth selectors, including local and global selectors, in order to make an adequate comparison. Therefore, we compare the above proposed local bootstrap selector with others that have been suggested for solving the estimation problem considered here. The global selectors evaluated were the plug-in selector and one based on crossvalidation (both proposed by Opsomer and Ruppert, 1998). In addition to the proposed bootstrap selector, we considered two local selectors: the theoretical optimum bandwidth defined in (18), and a local crossvalidation bandwidth selector which was defined by adapting it to the context that Vieu (1991) proposed for a unidimensional situation.

The local bootstrap bandwidth selector requires the choice of a pilot bandwidth parameter, $\mathbf{g}$. For this purpose, we can use a methodology analogous to that followed by González-Manteiga et al. (2004) when a linear local multivariate estimator is involved, and by Härdle and Marron (1991) in a simpler context. However, when backfitting estimators are involved, serious theoretical difficulties arise. Indeed, the asymptotics involved become extremely complex and treatable, meaningful expressions cannot be obtained. In our calculations we consider the pilot bandwidths obtained by a global crossvalidation selector.

4.1. Simulated additive models

We consider three additive models with component functions and design distributions specified as follows:

Model 1

$$m_1(x_1) = 1 - 6x_1 + 36x_1^2 - 53x_1^3 + 22x_1^5$$

and

$$m_2(x_2) = sin(5\pi x_2).$$

The explicative variables were generated from independent normal distributions with mean 0.5 and variance 1/9.

Model 2

$$m_1(x_1) = \frac{1}{2} sin(\pi x_1)$$

and

$$m_2(x_2) = sin(4\pi x_2).$$

The explicative variables are again considered to be independent but to be generated from a uniform distribution in the interval $[0, 1]$.

Model 3

$$m_1(x_1) = x_1^2$$

and

$$m_2(x_2) = x_2^2.$$

with the same design distribution as was defined for Model 2.

For all models, the residuals were generated from a distribution normal with mean zero and constant variance of 0.25. The sample size was $n = 100$, and 100 repetitions of each model were made.

The local linear smoother involved in the backfitting algorithm was calculated with a gaussian kernel, $K(x) = (2\pi)^{(-1/2)} \exp\left(-x^2/2\right)$. The resultant backfitting estimator was evaluated on a grid of 10×10 equally spaced estimation points (denoted by $ngrid$ the size of the grid).

4.2. Bootstrap approximation of MSE and bootstrap bandwidth selector

We now analyse the finite-sample behaviour of the bootstrap approximation, $\mathrm{MSE}^*(\mathbf{x}; \mathbf{h}, \mathbf{g})$, as a function of $\mathbf{x}$. For this purpose, we find the proximity between this bootstrap approximation and the theoretical surface, $\mathrm{MSE}(\mathbf{x}; \mathbf{h})$, over the grid of estimation points defined for the three models.

Thus we have implemented the exact expressions of the bootstrap bias and the bootstrap variance defined in (23) and (24), respectively, and the equivalent expression for (17) given by

$$\mathrm{MSE}(\mathbf{x};\mathbf{h}) = \mathrm{B}_{\mathbf{h}}(\mathbf{x})^2 + \mathrm{V}_{\mathbf{h}}(\mathbf{x}),$$

with bias and variance components defined in (2).

Figures 1, 2 and 3 represent the surfaces obtained for bootstrap MSE* and the theoretical MSE (both evaluated on their respective minimisers over $\mathbf{h}$, and the crossvalidation bandwidth for the pilot $\mathbf{g}$ involved in the bootstrap). These figures also show the corresponding bandwidth surfaces provided by the bootstrap selector, i.e., the components of $\mathbf{h}^*_{boot}(\cdot)$, and of the optimum bandwidth, $\mathbf{h}_{opt}(\cdot)$ (both being considered as functions of $\mathbf{x}$ and evaluated over the grid).

Note that the bootstrap and the theoretical surfaces are close in all cases, which points to the good performance of the bootstrap approximation. The surfaces become approximately overlapped at the interior of the estimation interval considered.

4.3. SiZer Map for exploring the behaviour of several bandwidth selectors

SiZer Map is a graphical tool for exploring the features of the data set which support the estimation curve problem. By using different colours (or different tones of grey for black and white versions), we can show the significant increase or decrease of the target curve, considering different smoothing levels. In this paper, SiZer allows us to evaluate the behaviour of several bandwidth selectors, by visualizing their position inside the colour space defined by this curve.

SiZer was first proposed by Chaudhuri and Marron (1999), in a univariate context, to determine the significance of the features of a target curve, such as peaks and valleys, by considering a family of smoothers, $\{\widehat{m}\,(x;h) : h \in [h_{min}, h_{max}]\}$. The procedure involves the construction of confidence intervals for the derivative, $m'\,(x;h)$, at the space defined by the smoothing parameter, h. Over the localization space defined by x and h, the figure shows the performance of the estimated curves by means of different colours: blue (black, for black and white versions) if the derivative is significantly positive; red (dark grey) is the derivative is significantly negative, and purple (light grey) if the derivative is not significantly nonzero. The lightest grey zones indicate smoothing parameter values that are too small and which invalidate the calculation of the estimated curve, i.e., the number of observations that fall within each local window is insufficient for estimation.

The confidence intervals for the derivative are computed by

$$\widehat{m}'\,(x;h) \pm q\sqrt{\widehat{\mathrm{Var}}\,(\widehat{m}'\,(x;h))},$$

where the quantile, q, is obtained by normal approximation or bootstrap techniques (see Chaudhuri and Marron (1999) for more details).

The range of bandwidths, $[h_{min}, h_{max}]$, can be defined in different ways, the usual method being to choose a range that is wide enough to show the

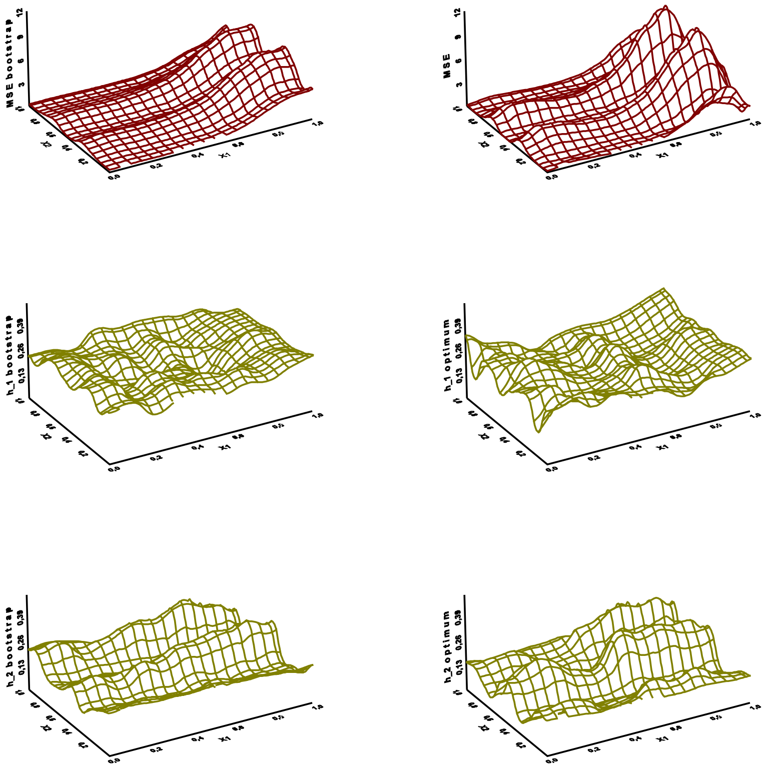

Figure 1: The top panels show the mean squared error surfaces for model 1: the bootstrap approximation (left) and the theoretical surface (right). The panels at the bottom represent the two components, bootstrap bandwidth (left) and optimal bandwidth (right).

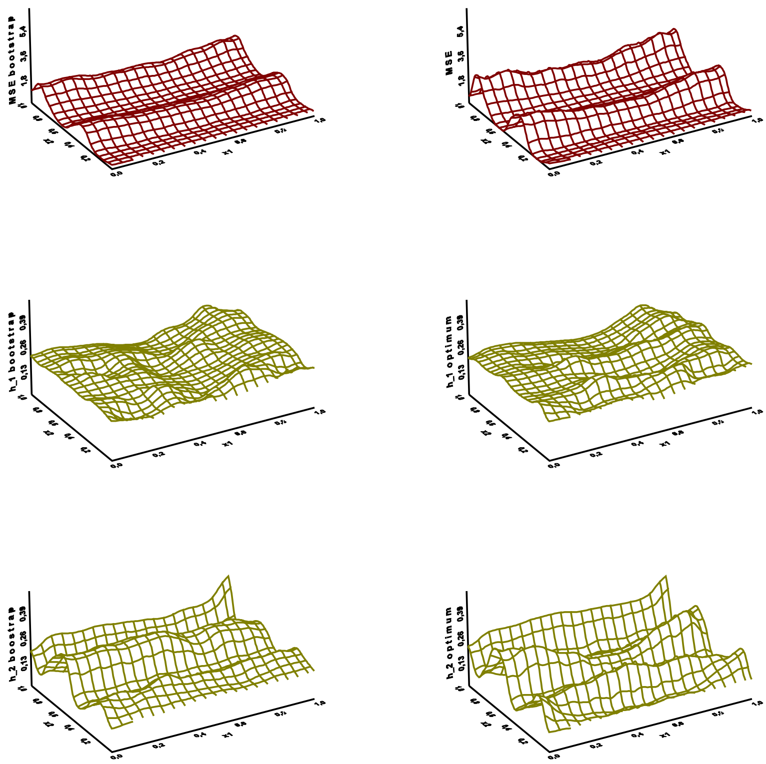

Figure 2: The top panels show the mean squared error surfaces for model 2: the bootstrap approximation (left) and theoretic surface (right). The panels at the bottom represent the two components, bootstrap bandwidth (left) and optimal bandwidth (right).

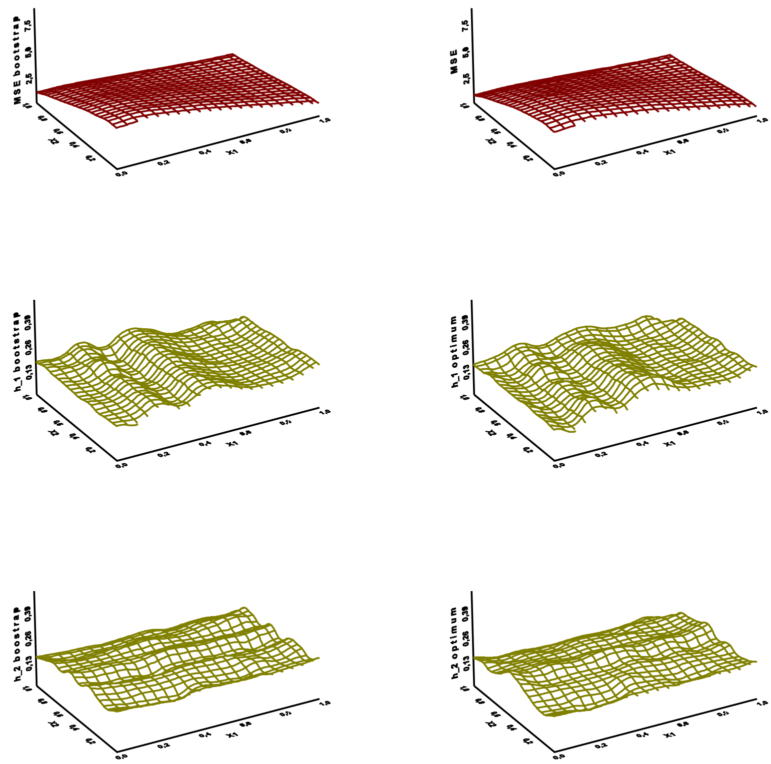

Figure 3: The top panels show the mean squared error surfaces for model 3: the bootstrap approximation (left) and theoretic surface (right). The panels at the bottom represent the two components, bootstrap bandwidth (left) and optimal bandwidth (right).

subjacent characteristics in each zone of the data set space, by using the family of smoothers.

Therefore, it is necessary to derive an expression for $\widehat{m}'(x;h)$ and also for $\widehat{\mathrm{Var}}(\widehat{m}'(x;h))$, so that they can be calculated in an efficient way (in computational terms) by speeding map construction. In this respect, the most usual procedure is to use wrapping techniques such as the binning method (Fan and Marron, 1994), whose main purpose is to reduce the number of kernel evaluations, on the basis of the similarity between many of these.

The extension of SiZer Map to a multidimensional context presents serious difficulties, even with regard to the graphical representation of the maps. One simple solution is to consider additive models, because these allow us to consider the effect of each covariate separately. Following this idea, an immediate extension to an additive model would be to construct as many SiZer Maps as there are covariates.

This reasoning inspired the work of Raya-Miranda et al. (2002), and is the basis of the exploratory study described in this paper. Let us now present some definitions and expressions included in this study.

Consider a family of backfitting estimators for an additive model such as the one considered in this paper, $\{\widehat{m}(\mathbf{x};\mathbf{h}) : h_d \in [h_{d,min}, h_{d;max}], 1 \leq d \leq D\}$, and define confidence intervals for the component functions, $\widehat{m}'_d(x_d;h_d)$. The dth curve shows the features of the m_d component by means of different colours, in a similar way to that adopted for the univariate context.

The confidence intervals for the derivative of the dth component are written as

$$\widehat{m}'_d(x_d;h_d) \pm \sqrt{\widehat{\mathrm{Var}}(\widehat{m}'_d(x_d;h_d))}.$$

Raya-Miranda et al. (2002) provide more details about the expressions of $\widehat{m}'_d(x_d;h_d)$ and $\widehat{\mathrm{Var}}(\widehat{m}'_d(x_d;h_d))$, and also for the binned approximations that are considered.

Figures 4 to 8 show the SiZer Maps for the three simulated additive models. Each figure consists of two types of curve, the so called Family Plot and the SiZer Map, both being constructed for each component in the model. Family Plot allows us to compare different choices of smoothing levels for estimating the components. Since separate figures are shown for each component function, the different curves in a component are given by changing its smoothing level, and so it is necessary to set the smoothing parameter value for the other component, or to define a Family Plot in a three dimensional space. In this paper, for reasons of simplicity, we have assumed the same smoothing level for the components, but trials were carried out to demonstrate the validity of such a restriction; in fact, other choices offered equivalent results (see figure 6).

The blue curves represent the estimates at different smoothing levels (the number of levels plotted can be chosen by the user, but the Matlab function `gpanal`, given by Marron (1997, 1998) considers eleven in a logarithmic scale). Besides these choices of the smoothing parameter, the figures include estimates using various, specific parameters; the black curve is associated to the local bootstrap selector proposed in this paper; the blue one represents the optimal theoretical bandwidth (also local); the red curve is the estimation with a local crossvalidation bandwidth; green shows the global plug-in selector, and yellow is used to plot estimations with a global crossvalidation bandwidth.

In the figure described as SiZer Map, the horizontal axis represents the range of the given covariate, and the smoothing levels are shown on the vertical axis (in a logarithmic scale). Here, we have considered the same smoothing level for the two additive components. The figure shows curves associated to particular choices of bandwidths, both local and global. The solid black line represents the plug-in selector and the dashed one shows a global crossvalidation bandwidth. The solid white curve is the local bootstrap bandwidth, the white dotted curve is associated to the optimal theoretical bandwidth (local), and the white dashed curve represents a local crossvalidation bandwidth (this selector has been evaluated only for model 1 because of the considerable computational time needed for its calculation; trials were carried out for the other models, showing a similar behaviour).

The 100 replicas carried out on each model would provide 100 SiZer Maps. In order to summarize them by eliminating the sample effect, the following procedure has been established: assign the colours for the maps at each localization point when the conditions defining these colours (defined before) occur for at least a given percentage of replicas. By following this idea, we considered several percentage values to obtain an adequate comparison and to draw useful conclusions about the problem of how to replicate the maps. Here, we include two percentages for model 1, i.e., 75% (figure 4) and 95% (figure 5). The appearance of the maps (for both the percentages and the other values tested) was similar for purposes of exploration and so we only include the 75% for the remaining simulated models (figures 7 and 8).

Figure 6 shows the SiZer Map for model 1, with the particular feature that the plug-in bandwidth is used for the other component (not plotted on the map). This figure shows that the restriction of taking the same smoothing level for both components does not hide any of the information provided by the maps.

All of the maps show that the selectors evaluated, both local and global methods, move with the localization falling inside the resolution levels, thus enabling us to detect the most significant characteristics of the estimated components. The greatest differences can be observed when components with a high degree of curvature are estimated. In these cases, the local optimum moves from the lowest resolution levels at the peaks and in the valleys, to higher levels at more linear zones of the components. This also happens at boundaries, probably in an attempt to obtain enough observations for the purpose of estimation. The local bootstrap selector moves with localizations located very close to the optimum, but local crossvalidation adapts to data by choosing very low smoothing levels.

Models with low curvature, as is the case in model 3, can be satisfactorily estimated by considering global selectors, such as the plug-in selector considered here, or the crossvalidation selector. Here, SiZer Map shows the proximity of all the selectors (both global and local) but the small variations, which the local bootstrap and optimum describe, provide more irregular estimates, as shown in the following section.

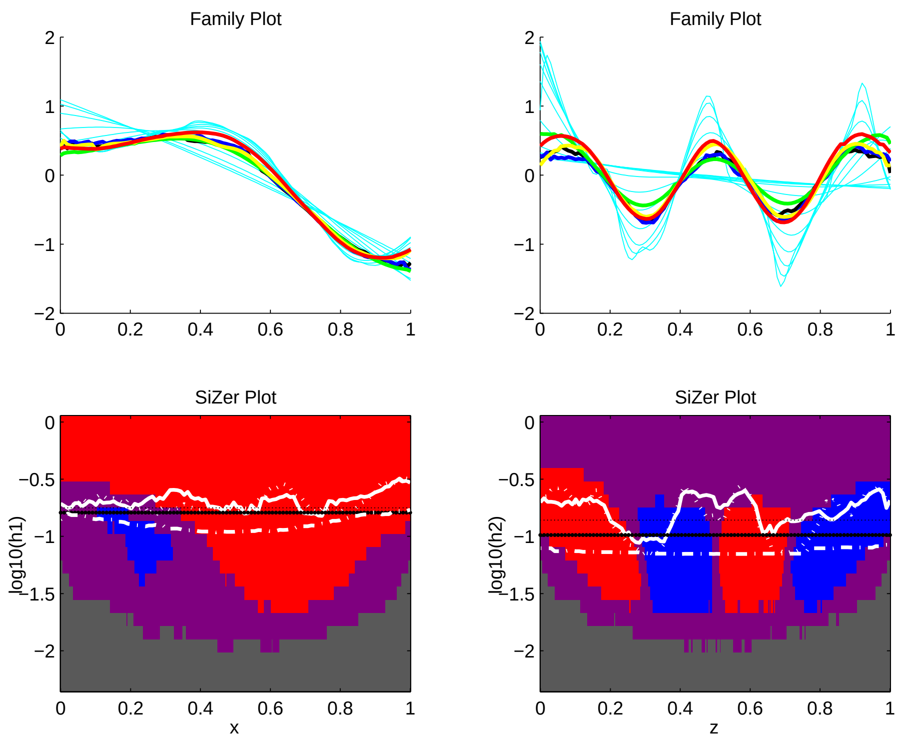

Figure 4: SiZer Map for model 1. The colours were chosen by checking the usual conditions for at least 75% of the replicas.

4.4. Estimated regression surfaces

In this section we evaluate the performance of the additive backfitting estimates with local bootstrap bandwidths. For this purpose, we make comparisons with other well-known selectors by using a measure of the estimation error versus the estimation interval. The calculated measure was the so called Integrated Squared Error (ISE), which is defined for each replica (r), by

$$ISE\,(r) = \frac{1}{ngrid} \sum_{k=1}^{ngrid} \left(\widehat{m}\left(\mathbf{z}_k\right) - m\left(\mathbf{z}_k\right)\right)^2,$$

where $\{\mathbf{z}_k : \ k = 1, \ldots, ngrid\}$ denotes the grid of estimation points used.

Figure 9 shows the surfaces estimated for each model simulated. The densities of ISEs (computed for the estimates with all the bandwidth selectors considered and the three simulated models) are plotted in figure 10.

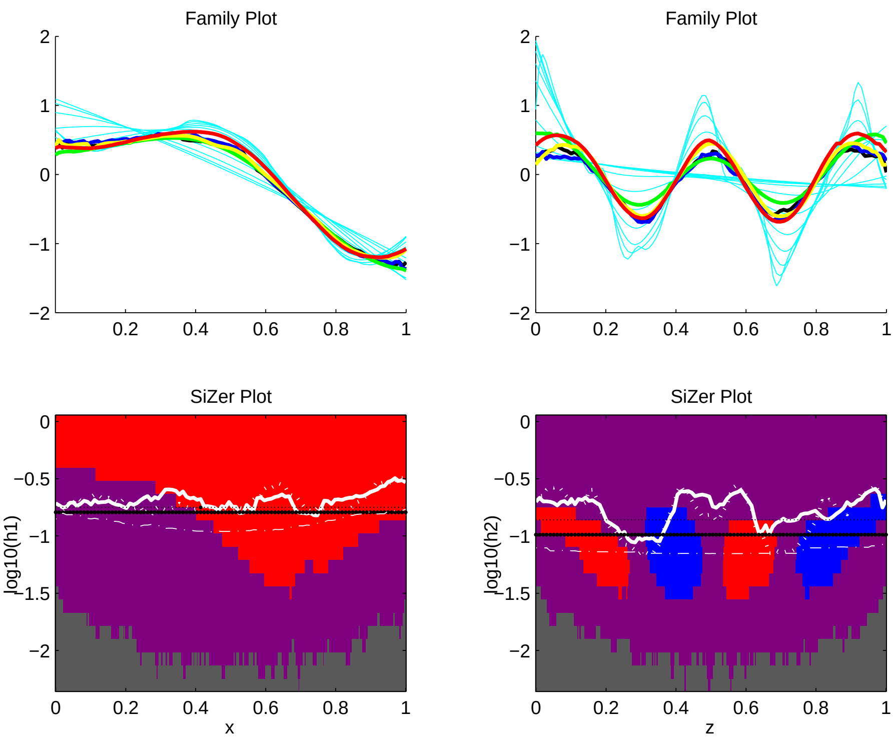

Figure 5: SiZer Map for model 1. The colors have been chosen by checking the usual conditions for at least 95% of the replicas.

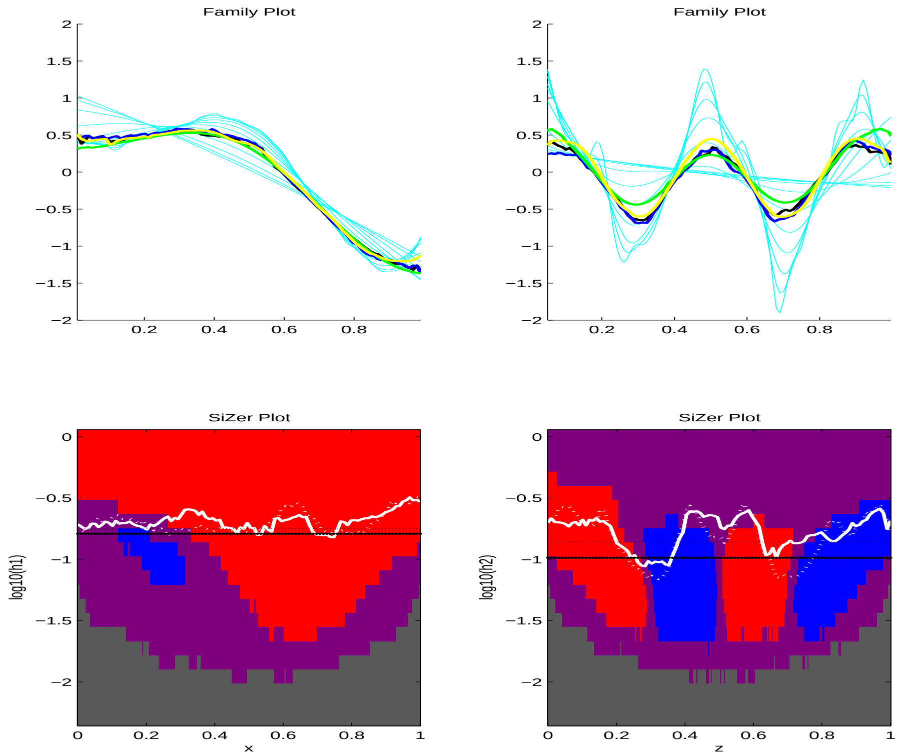

Figure 6: SiZer Map for model 1 with a fixed (plug-in) bandwidth for the other (non plotted) component. The colours were chosen by checking the usual conditions for at least 75% of the replicas.

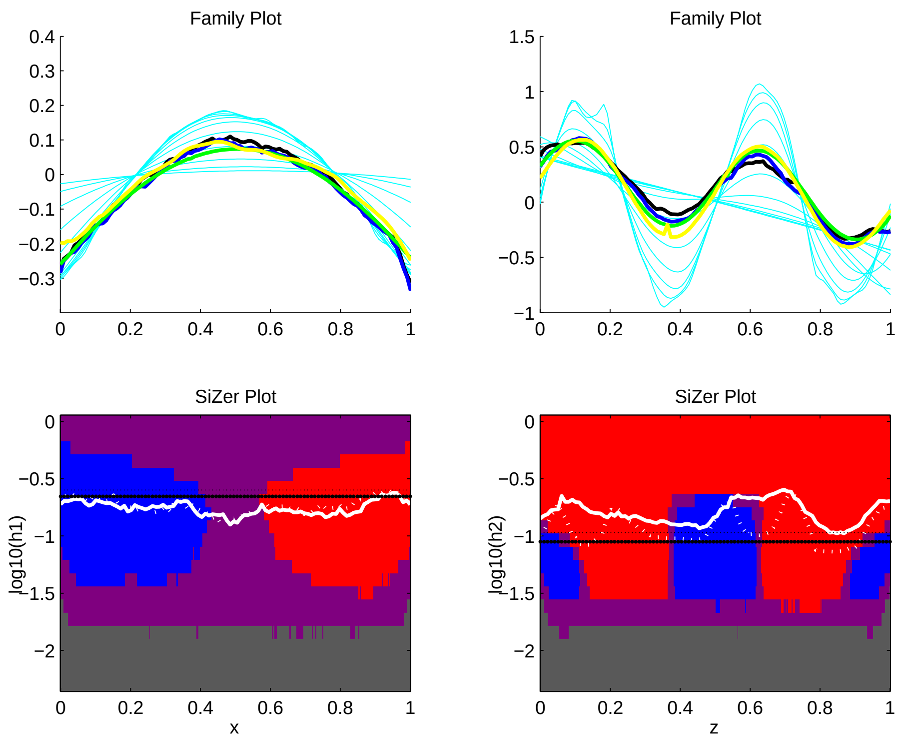

Figure 7: SiZer Map for model 2. The colours were chosen by checking the usual conditions for at least 75% of the replicas.

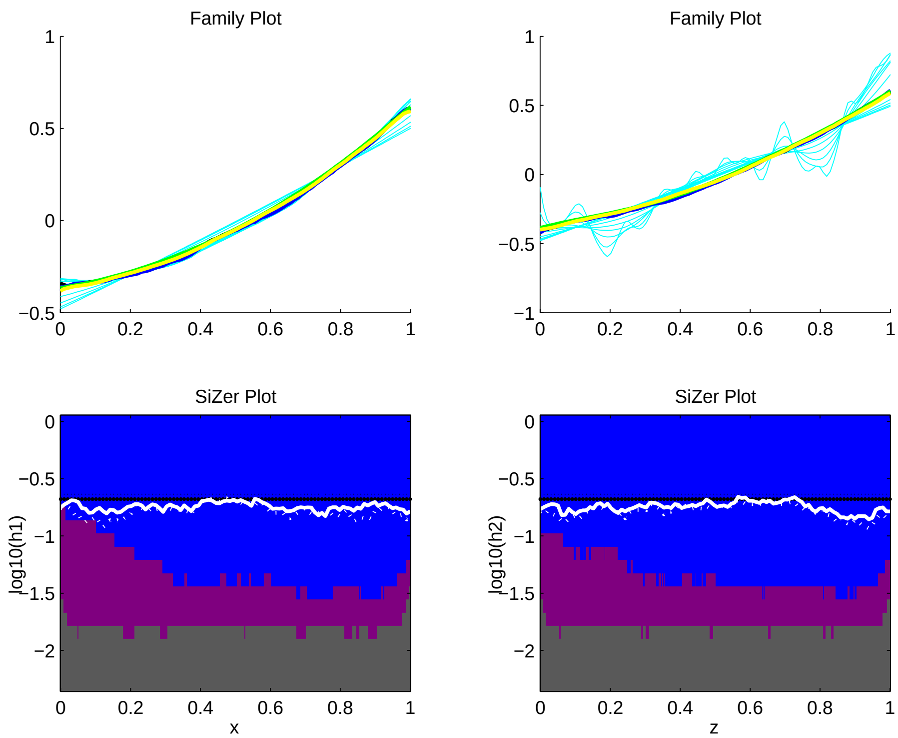

Figure 8: SiZer Map for model 3. The colours were chosen by checking the usual conditions for at least 75% of the replicas.

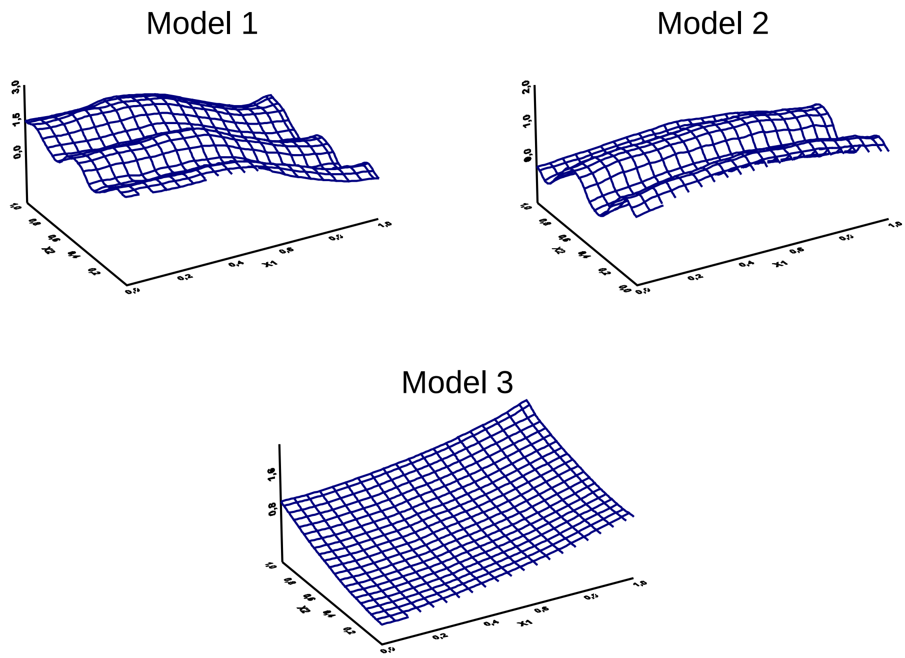

Figure 9: Estimated regression surfaces with local bootstrap bandwidth parameters.

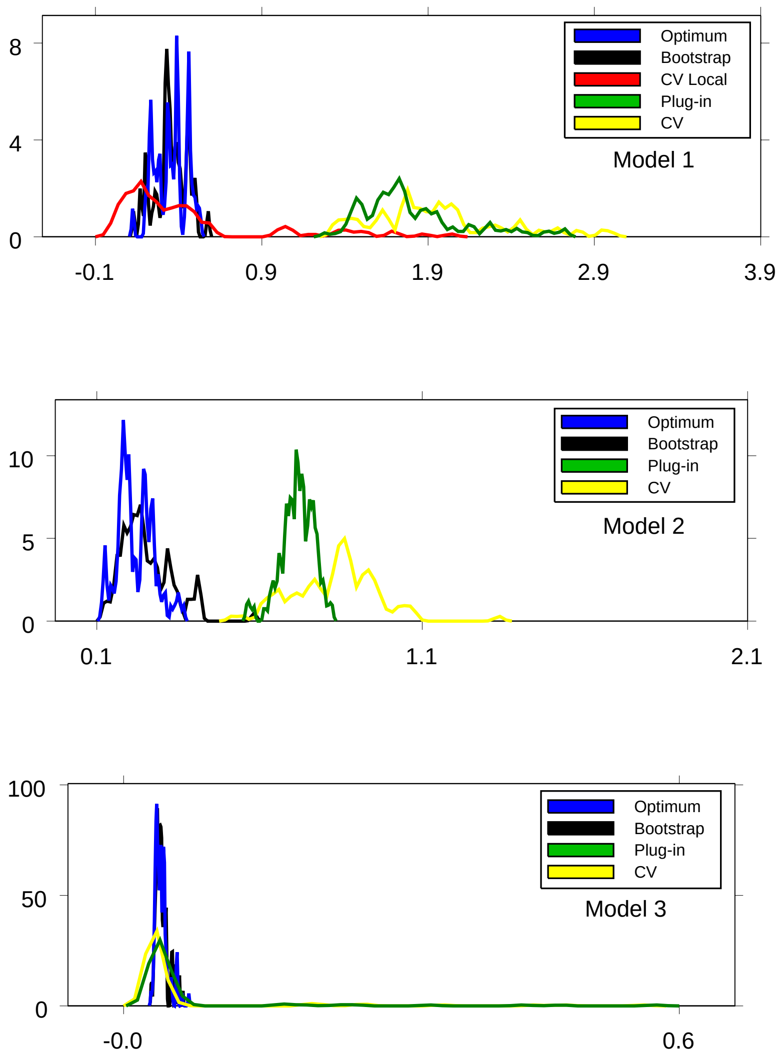

Figure 10: ISE density. For each model, the curves represent this measure when each of the bandwidth selectors compared are used.

Note the proximity of the ISE curves for the local bootstrap bandwidth and the optimal theoretical bandwidth, for all the models simulated. Indeed, these local selectors provide additive estimations which perform considerably better than those associated to global bandwidths, like the one obtained using plug-in and crossvalidation selectors, when models 1 and 2 are considered. For model 3, the behaviour of the selectors is similar but features the large tails associated to the global selectors (thus leading to large errors). This also happens with the local crossvalidation selector evaluated on model 1.

Also shown are the box-plots associated to ISE values for each selector, see figure 11. Here we can observe the high variability of local crossvalidation in the replicas (as expected, because of the known low convergence rate), in comparison with the homogeneous values provided by local bootstrap for model 1. Again, it is interesting to observe the behaviour of global selectors compared with the local ones for model 3, this difference being clearly apparent in the figure. Of course, the low curvature of the regression surface yields benefits when what is considered is a constant smoothing level with many outliers (which do not appear associated to the local selectors).

4.5. Discussion

The empirical study performed reveals the good behaviour of the wild bootstrap methodology for bandwidth selection purposes when backfitting estimators are considered. The proposed bootstrap estimates are easy to implement and the local bandwidth selector that is derived does not require lengthy computational times, in comparison with other local versions such as the local crossvalidation tested here. Indeed, the crossvalidation selector requires a high number of estimates and the calculation process becomes very slow because of the backfitting algorithm involved. Besides these computational problems, the crossvalidation selector at some localizations chooses bandwidths that are too small and which produce high variability in the estimated regression surface; moreover, in some cases it is impossible to compute estimates. Analogous similarities and problems were pointed out by González-Manteiga et al. (2004) for the case of the multivariate local linear smoother. The superiority of local selectors over global selectors in general is again apparent, moreover, when the data describe surfaces with a high degree of curvature.

SiZer Map provides an intuitive space which is adequate for interpreting and visualizing the performance of bandwidth selectors. As Chaudhuri and Marron (1999) observed, local selectors provide estimates which adapt very well to data and can capture the most significant features of the subjacent surfaces. As a general conclusion, all the maps show that the proposed bootstrap selector closely follows the local optimum, while local crossvalidation locates always below, at undersmoothing levels.

5. CONCLUDING REMARKS

We propose a new bandwidth selector for nonparametric additive modelling, one that is based on bootstrap estimates of the conditional mean squared error and

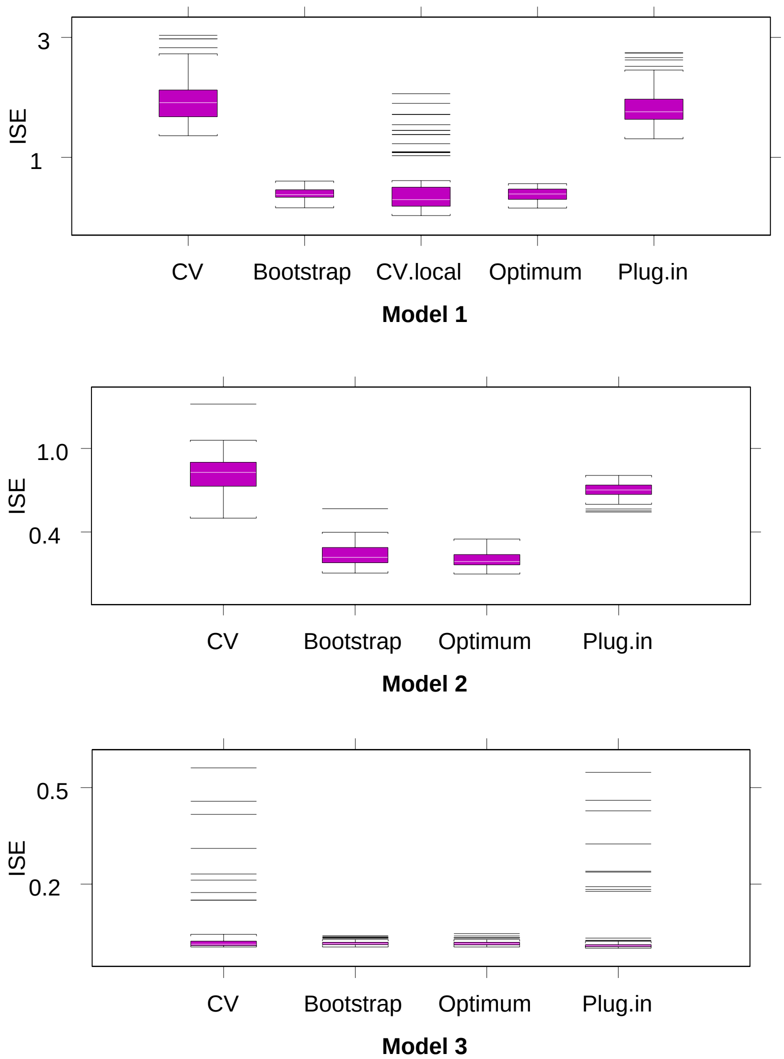

Figure 11: ISE box plots for each model simulated. The boxes represent this error measure when each of the bandwidth selectors compared are used.

varies with the localizations of surfaces. The selector does not require complex asymptotic expressions of the error criterion (the conditional MSE) and is not derived by assuming strict conditions for the model and the sample sizes, as is the case with those selectors based on plug-in methodologies. In addition, the bootstrap selector can be easily evaluated in practice for additive estimation methods, including bivariate backfitting, general D-variate marginal integration and efficient mixed methods such as the one proposed recently by Kim et al. (1999). The difficulties in extending the selector to D-variate ($D > 2$) backfitting arise from the theoretical difficulties pointed out by Opsomer and Ruppert (1997), who, for such a situation, proposed a plug-in selector, assuming independent covariates. The bootstrap estimates are calculated without involving Monte Carlo approximations, and so the bootstrap selector has a low computational cost in comparison with other local methods like crossvalidation.

The local selector produces a good imitation of the local optimum, as shown in the empirical studies described in this paper, and is a serious competitor to other selectors proposed, including those based on plug-in methodologies and the classical selectors based on crossvalidation (both global and local versions). Comparisons made with SiZer Map reveal interesting characteristics about the behaviour of local selectors versus the global ones. Global and local selectors provide different patterns of performance, depending on the most significant features of the surfaces to be detected by the estimation method in question.

APPENDIX: SKETCH OF PROOFS

Proof of Lemma 1. From expressions (8) and (9), by substituting the matrices $\mathbf{W}_1, \mathbf{W}_2$ given by (6), the additive components $\widehat{m}_1(x_1)$ and $\widehat{m}_2(x_2)$ can be written as:

$$\widehat{m}_1(x_1) = s_{1,x_1}^T (I - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1}((I - \mathbf{S}_2^*)\mathbf{Y} \quad \text{and}$$
$$\widehat{m}_2(x_2) = s_{2,x_2}^T (I - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1}((I - \mathbf{S}_1^*)\mathbf{Y}.$$

By considering asymptotic representations of the before involved inverses as well as matrices $\mathbf{S}_1^*$ and $\mathbf{S}_2^*$, by following the lemmas 3.1 and 3.2 in Opsomer and Ruppert (1997), it follows:

$$\begin{aligned} \widehat{m}_1(x_1) \quad &= s_{1,x_1}^T \left[\mathbf{I} + O(\mathbf{1}\mathbf{1}^T/n\right] (\mathbf{I} - \mathbf{S}_2 + \mathbf{1}\mathbf{1}^T/n + o(\mathbf{1}\mathbf{1}^t/n))\mathbf{Y} \\ &\equiv \mathbf{w}_1(x_1)\mathbf{Y}. \end{aligned} \tag{25}$$

Again, using asymptotic representations but for elements of $\mathbf{S}_2$ and s_{1,x_1}, first the elements of the matrix $\mathbf{I} - \mathbf{S}_2 + \mathbf{1}\mathbf{1}^T/n$ can be expressed as:

$$\begin{aligned} &1 - n^{-1} f_2^{-1}(X_{i2}) K_{h_2}(0) + n^{-1} + o_p(n^{-1}h_2^{-1}) \quad \text{if } i = j \;\; \text{and} \\ &1 - n^{-1} f_2^{-1}(X_{i2}) K_{h_2}(0) + n^{-1} + o_p(n^{-1}h_2^{-1}) \quad \text{if } i \neq j, \end{aligned}$$

for each $i, j = 1, \ldots, n$. And by simple calculations the elements of $\mathbf{w}_1(x_1)$ (25) can be approximated (here it's been used the latest approximation in assumption A2) by

$$n^{-1} f_1^{-1}(x_1) n^{-1} K_{h_1}(X_{i1} - x_1) + o_p(n^{-1}h_1^{-1}) + o_p(n^{-1}h_2^{-1}h_1^{-1}) - o_p(1) + o(n^{-1})$$

for $1 \leq i \leq n$. Analogous calculations for the second component complete the proof of the lemma. □

Proof of Theorem 1. The proof follows analogous steps to proof of Theorem 1 in González-Manteiga et al. (2004). Thus, we consider the usual decomposition of MSE in (17) into bias term, $\mathrm{B}_{\mathbf{h}}(\mathbf{x})$ and a variance term, $\mathrm{V}_{\mathbf{h}}(\mathbf{x})$, and also for the bootstrap approximation, the bootstrap bias, $\mathrm{B}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x})$ in (23) and the bootstrap variance, $\mathrm{V}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x})$ in (24). We'll prove that the difference between each term and its bootstrap approximation tends to zero in probability:

Consider first the difference between biases, the consistency of backfitting estimators gives that this difference is asymptotically equivalent to the sum of two terms which are denoted and defined by

$$T_{1,1} = \sum_{i=1}^{n} w_i(\mathbf{x},\mathbf{h})\, w_i(\mathbf{X}_i,\mathbf{g})\,(m(\mathbf{X}_i) - Y_i) \tag{26}$$

$$T_{1,2} = \sum_{i \neq j} w_i(\mathbf{x},\mathbf{h})\, w_j(\mathbf{X}_i,\mathbf{g})\,(m(\mathbf{X}_i) - Y_j) \tag{27}$$

with $w_i(\cdot,\cdot)$ defined in (10). Lemma 1 gives the following asymptotic approximation for them:

$$w_i(\mathbf{x},\mathbf{h}) \approx n^{-1} f_1^{-1}(x_1)\, K_{h_1}(X_{i1} - x_1) + n^{-1} f_2^{-1}(x_2)\, K_{h_2}(X_{i2} - x_2) \tag{28}$$

which yields equivalent expressions for terms (26) and (27), which will be denoted by $\widetilde{T}_{1,1}$ and $\widetilde{T}_{1,2}$, respectively.

Now, since term $\widetilde{T}_{1,1}$ has mean zero and variance

$$O\left(n^{-3}(h_1 h_2^{-1} + h_1^{-1} h_2)^2 (g_1^{-1} + g_2^{-1})^2\right),$$

by applying Markov's inequality

$$\widetilde{T}_{1,1} = O_p\left(n^{-3/2}(h_1 h_2^{-1} + h_1^{-1} h_2)(g_1^{-1} + g_2^{-1})\right).$$

Similarly for second term it's been computed its expectation, yielding $o(1)$, and its variance being $O\left(n^{-2}(h_1^{-1} h_2 + h_1 h_2^{-1})(g_1^{-1} g_2 + g_1 g_2^{-1})\right)$.

Again Markov's inequality gives that

$$\widetilde{T}_{1,2} = O_p\left(n^{-1}(h_1^{-1} h_2 + h_1 h_2^{-1})^{1/2}(g_1^{-1} g_2 + g_1 g_2^{-1})^{1/2}\right).$$

Now consider the difference between variances:

$$\begin{aligned}
\mathrm{V}^*_{\mathbf{h},\mathbf{g}}(\mathbf{x}) - \mathrm{V}_{\mathbf{h}}(\mathbf{x}) =& \sum_{i=1}^{n} w_i(\mathbf{x},\mathbf{h})^2 \left(\varepsilon_i^2 - \sigma^2(\mathbf{X}_i)\right) + \\
&+ \sum_{i=1}^{n} w_i(\mathbf{x},\mathbf{h})^2 \left(\widehat{m}_{\mathbf{g}}(\mathbf{X}_i) - m(\mathbf{X}_i)\right)^2 - \\
&- 2\sum_{i=1}^{n} w_i(\mathbf{x},\mathbf{h})^2 \varepsilon_i \left(\widehat{m}_{\mathbf{g}}(\mathbf{X}_i) - m(\mathbf{X}_i)\right) = \\
=& \vartheta_1 + \vartheta_2 - 2\vartheta_3,
\end{aligned}$$

and operate separately on each term ϑ_i, $i = 1, 2, 3$ in a similar way that used before with the biases. We prove that the expectation of the first term vanishes and the variance is $O(n^{-3}(h_1^{-1} + h_2^{-1})^4 h_1 h_2)$, therefore $\vartheta_1 = O_p(n^{-3/2}(h_1^{-1} + h_2^{-1})^2 (h_1 h_2)^{1/2})$.

By developing the square in the second term, it can be separate into a sum of the terms (an U-statistics with order two and other with orden three) given by:

$$\begin{aligned}\vartheta_{2,1} &= \sum_{i=1}^{n} w_i (\mathbf{x}, \mathbf{h})^2 w_i (\mathbf{X}_i, \mathbf{g})^2 (Y_i - m(\mathbf{X}_i))^2 + \\ &+ \sum_{i \neq j} w_i (\mathbf{x}, \mathbf{h})^2 w_i (\mathbf{X}_i, \mathbf{g})^2 (Y_j - m(\mathbf{X}_i))^2\end{aligned}$$

and

$$\vartheta_{2,2} = \sum_{i=1}^{n} w_i (\mathbf{x}, \mathbf{h})^2 \sum_{\substack{j \neq l \\ l, j \neq i}} w_j (\mathbf{X}_i, \mathbf{g}) w_l (\mathbf{X}_i, \mathbf{g}) (Y_j - m(\mathbf{X}_i)) (Y_l - m(\mathbf{X}_i)).$$

Analogous arguments and similar calculations to the before performed yield that

$$\begin{aligned}\vartheta_2 = \; & O_p\left(n^{-5/2} \left(h_1^{-1} + h_2^{-1}\right)^2 \left(g_1^{-1} + g_2^{-1}\right)^2 h_1 h_2\right) + \\ & + O_p\left(n^{-7/2} \left(h_1^{-1} + h_2^{-1}\right)^2 \left(g_1^{-1} + g_2^{-1}\right)^2 (h_1 h_2)^{1/2}\right).\end{aligned}$$

Finally operate on third term and obtain that

$$\begin{aligned}\vartheta_3 = \; & O_p\left(n^{-5/2} \left(h_1^{-1} + h_2^{-1}\right)^2 \left(g_1^{-1} + g_2^{-1}\right) (h_1 h_2)^{1/2}\right) + \\ & + O_p\left(n^{-2} \left(h_1^{-1} + h_2^{-1}\right)^2 \left(g_1^{-1} + g_2^{-1}\right) (h_1 h_2 g_1 g_2)^{1/2}\right),\end{aligned}$$

completing the proof of theorem.□

References

[1] Buja, A., Hastie T. and Tibshirani, R. (1989), "Linear smoothers and additive models" (with discussion), *The Annals of Statistics*, 17, 453–555.

[2] Chaudhuri, P. and Marron, J.S. (1999), "SiZer for exploration of structures in curves", *Journal of the American Statistical Association*, 94, 807–823.

[3] Dalelane, C. (1999), "Bootstrap confidence bands for the integration estimator in additive models," Ph.D. dissertation, Faculty of Mathematics, Humboldt University of Berlin.

[4] Fan, J. and Gijbels, I. (1995), "Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation", *Journal of the Royal Statistics Society*, Ser.B, 57, 371–354.

[5] Fan, J. and Marron, J.S. (1994), "Fast implementation of nonparametric curve estimators", *Journal of Computational and Graphical Statistics*, 3, 35–56.

[6] Friedman, J.H. and Stuetzle, W. (1981), "Projection pursuit regression", *Journal of the American Statistical Association*, 76, 817–823.

[7] González-Manteiga, W., Martínez-Miranda, M.D. and Pérez-González, A. (2004), "The choice of smoothing parameter in nonparametric regression through wild bootstrap", *Computational Statistics and Data Analysis*, 47, 487–515.

[8] Gu, C. and Wahba, G. (1988), "Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method", *SIAM Journal on Scientific and Statistical Computing*, 12, 383–398.

[9] Härdle, W., Hall, P. and Marron, J.S. (1988), "How far are automatically chosen regression smoothing parameters from their optimum?", *Journal of the American Statistical Association*, 83, 86–95.

[10] Härdle, W., Huet, S., Mammen, E. and Sperlich, S. (2001), "Semiparametric additive indices for binary response and generalized additive models", Discussion Paper 95, Sonderforschungsbereich 373.

[11] Härdle, W. and Marron, J. S. (1991), "Bootstrap simultaneous error bars for nonparametric regression", *The Annals of Statistics*, 19, 778–796.

[12] Hastie, T.J. and Tibshirani, R. (1990), *Generalized additive models*, Washington, D.C. Chapman & Hall.

[13] Hengartner, N.W. (1996), "Rate Optimal Estimation of Additive Regression via the Integration Method in Presence of Many Covariates", Preprint, Department of Statistics, Yale University, `http:\\www.stat.yale.edu.`

[14] Kauermann, G. and Opsomer, J.D. (2003), "Local likelihood estimation in generalized additive models", *Scandinavian Journal of Statistics*, 30, 317–337.

[15] Kauermann, G. and Opsomer, J.D. (2004), "Generalized cross-validation for bandwidth selection of backfitting estimates in generalized additive models", *Journal of Computational and Graphical Statistics*, 13, 66–89.

[16] Kim, W., Linton, O.B. and Hengartner, N.W. (1999), "A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals", *Journal of Computational and Graphical Statistics*, 8, 278–297.

[17] Linton, O.B. (1997), "Efficient estimation of additive nonparametric regression models", *Biometrika*, 84, 469–473.

[18] Linton, O.B. and Nielsen, J.P. (1995), "A kernel method of estimating structured nonparametric regression based on marginal integration", *Biometrika*, 82, 93–100.

[19] Opsomer, J.D. (2000), “Asymptotic properties of backfitting estimators”, *Journal of Multivariate Analysis*, 75, 166–179.

[20] Opsomer, J.D. and Ruppert, D. (1997), “Fitting a bivariate additive model by local polynomial regression”, *The Annals of Statistics*, 25, 186–211.

[21] Opsomer, J.D. and Ruppert, D. (1998), “A fully automated bandwidth selection method for fitting additive models”, *Journal of the American Statistical Association*, 93, 605–619.

[22] Opsomer, J.D. and Ruppert, D. (1999), “A Root-n consistent backfitting estimator for semiparametric additive modeling”, *Journal of Computational and Graphical Statistics*, 8, 715–732.

[23] Raya-Miranda, R., Martínez-Miranda, M.D. and González-Carmona, A. (2002), “Exploring the structure of regression surfaces by using SiZer Map for additive models”, in *Proceedings in Computational Statistics, COMPSTAT 2002*, eds. W. Härdle and B. Rönz, Humboldt University of Berlin, Berlin.

[24] Ruppert, D. (1997), “Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation”, *Journal of the American Statistical Association*, 92, 1049–1062.

[25] Severance-Lossin, E. and Sperlich, S. (1997), “Estimation of derivatives for additive separable models”, Discussion paper, SFB 373. Humboldt-University, Berlin.

[26] Sperlich, S., Tjϕstheim, D. and Yang, L. (2002), “Nonparametric estimation and testing of interaction in additive models”, *Economic Theory*, 18, 197–251.

[27] Sperlich, S., Linton, O.B. and Härdle, W. (1999), “Integration and backfitting methods in additive models-finite sample properties and comparison”, *Test*, 8, 419–458.

[28] Stone, C.J. (1985), “Additive regression and other nonparametric models”, *The Annals of Statistics*, 13, 689–705.

[29] Stone, C.J. (1986), “The dimensionality reduction principle for generalized additive models”, *The Annals of Statistics*, 14, 590–606.

[30] Vieu, P. (1991), “Nonparametric regression: Optimal local bandwidth choice”, *Journal of the Royal Statistical Society*, Ser. B, 53, 453–464.

[31] Vieu, P. (1993), “Bandwidth selection for kernel regression: a survey”, in *Computer intensive methods in statistics*, Statistics and Computing (Vol. 1, pp. 134-149), eds. W. Hardlë and L. Simar, Berlin.

[32] Wu, C.F.J. (1986), “Jackknife, bootstrap and other resampling methods in regression analysis”, *The Annals of Statistics*, 14, 1261–1350.

[33] Yang, L., Sperlich, W. and Härdle, W. (2003), "Derivative estimation and testing in generalized additive models", *Journal of Statistical Planning and Inference*, 115, 521–542.