Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Partly linear models on Riemannian manifolds

Wenceslao Gonzalez-Manteiga [a], Guillermo Henry [b] & Daniela Rodriguez [b]

[a] Departmento de Estadistica e Investigacion Operativa, Universidad de Santiago de Compostela, Santiago de Compostela, Spain

[b] Departmento de Matematica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, CABA, Argentina

Available online: 01 May 2012

To cite this article: Wenceslao Gonzalez-Manteiga, Guillermo Henry & Daniela Rodriguez (2012): Partly linear models on Riemannian manifolds, Journal of Applied Statistics, DOI:10.1080/02664763.2012.683169

To link to this article: http://dx.doi.org/10.1080/02664763.2012.683169

Journal of Applied Statistics
2012, iFirst article

Partly linear models on Riemannian manifolds

Wenceslao Gonzalez-Manteiga[a], Guillermo Henry[b] and Daniela Rodriguez[b]*

[a]*Departmento de Estadistica e Investigacion Operativa, Universidad de Santiago de Compostela, Santiago de Compostela, Spain;* [b]*Departmento de Matematica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, CABA, Argentina*

(*Received 18 April 2011; final version received 3 April 2012*)

In partly linear models, the dependence of the response y on $(\mathbf{x}^{\mathrm{T}}, t)$ is modeled through the relationship $y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + g(t) + \varepsilon$, where ε is independent of $(\mathbf{x}^{\mathrm{T}}, t)$. We are interested in developing an estimation procedure that allows us to combine the flexibility of the partly linear models, studied by several authors, but including some variables that belong to a non-Euclidean space. The motivating application of this paper deals with the explanation of the atmospheric SO_2 pollution incidents using these models when some of the predictive variables belong in a cylinder. In this paper, the estimators of $\boldsymbol{\beta}$ and g are constructed when the explanatory variables t take values on a Riemannian manifold and the asymptotic properties of the proposed estimators are obtained under suitable conditions. We illustrate the use of this estimation approach using an environmental data set and we explore the performance of the estimators through a simulation study.

Keywords: environmental data; hypothesis test; non-parametric estimation; partly linear models; Riemannian manifolds

1. Introduction

Partly linear models were introduced by Engle *et al.* [6] to analyze the relationship between electricity usage and average daily temperature. In recent years, these model have gained a lot of attention for their use in the exploration of the nature of complex nonlinear phenomena. Partly linear models have been widely studied in the literature, see, for example, [1,4,17], among others. These models allow us to model the response variable with a set of predictors that enter into the models linearly while one of them is considered non-parametrically in the models.

In this paper, we discuss the application of these models to the sulfur dioxide (SO_2) pollution problem. More specifically, we are interested in modeling the emission of SO_2 through variables such as the temperature and the direction and speed of the wind. It is important to mention that the variables of wind have a common structure. Also, the circular structure of the wind direction allows us to consider a cylinder as the space where the speed and direction take values.

*Corresponding author. Email: drodrig@dm.uba.ar

ISSN 0266-4763 print/ISSN 1360-0532 online

http://dx.doi.org/10.1080/02664763.2012.683169
http://www.tandfonline.com

However, partly linear models do not seem to include this structure. The case studied in this paper is only an example of predictive variables taking values on a Riemannian manifold rather than on an Euclidean space. Some other examples of variables taking values on a non-Euclidean space can be found in meteorology, astronomy, geology and other fields, which include natural distributions on spheres, tangent bundles and Lie groups. Research on the statistical analysis of variables with one of these structures was done by Bhattacharya and Patrangenaru [3] and Mardia [13] and, more recently, by Hendriks and Landsman [9], Henry and Rodriguez [10], Pelletier [14] and Pennec [15].

The aim of this work is to study partly linear models when the explanatory variable t takes values on a Riemannian manifold, that is, when the variable to be modeled in a non-parametric way is in a manifold. We introduce an estimation procedure that includes this structure of the variables.

The paper is organized as follows. In Section 2, we construct estimates for these models and give a review of the non-parametric estimation on Riemannian manifolds proposed in [14]. In Section 3, we present the asymptotic behavior of the proposed estimators under regular assumptions. In Section 4, we explore the performance of the estimators through a simulation study and we give an example using real data. Moreover, we review a cross-validation procedure for partly linear models. The proofs of the theoretical results presented in Section 3 are given in Appendix 1.

2. Estimators

2.1 *Model and estimators*

Let $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)$ be an iid random vector valued in $\mathbb{R}^{p+1} \times M$ with a distribution identical to that of $(y, \mathbf{x}^{\mathrm{T}}, t)$, where (M, ξ) is a Riemannian manifold of dimension d. Partly linear models assume that the relation between the response variable y_i and the covariates $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ can be represented as

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + g(t_i) + \varepsilon_i, \quad 1 \le i \le n, \tag{1}$$

where the errors ε_i are independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}$ and also of $E(\varepsilon_i|\mathbf{x}_i, t_i) = 0$. In many situations, it seems reasonable to suppose that a relationship between the covariates $\mathbf{x}$ and t exists; so as has been done in [1,17], we assume that for $1 \le j \le p$,

$$x_{ij} = \phi_j(t_i) + \eta_{ij}, \quad 1 \le i \le n, \tag{2}$$

where the errors η_{ij} are independent. Denote $\phi_0(\tau) = E(y|t = \tau)$ and $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))$, then we have that $g(t) = \phi_0(t) - \boldsymbol{\phi}(t)^{\mathrm{T}}\boldsymbol{\beta}$ and hence $y - \phi_0(t) = (\mathbf{x} - \boldsymbol{\phi}(t))^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$. These equations suggest estimating the unknown functions and parameters as follows. Let $\hat{\phi}_j(t)$ be the non-parametric estimators of ϕ_j, $0 \le j \le p$. Regarding the estimation of the parameter $\boldsymbol{\beta}$, we note that using the non-parametric estimators of the functions ϕ_j, the regression parameter can be estimated considering the least-square estimators obtained by minimizing

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} [(y_i - \hat{\phi}_0(t_i)) - (\mathbf{x}_i - \hat{\boldsymbol{\phi}}(t_i))^{\mathrm{T}}\boldsymbol{\beta}]^2,$$

where $\hat{\boldsymbol{\phi}}(t) = (\hat{\phi}_1(t), \ldots, \hat{\phi}_p(t))$. Then, the function g can be estimated as $\hat{g}(t) = \hat{\phi}_0(t) - \hat{\boldsymbol{\phi}}(t)^{\mathrm{T}}\hat{\boldsymbol{\beta}}$.

Note that the regression functions correspond to the predictors taking values in a Riemannian manifold; non-parametric kernel-type estimators adapted to this structure have been considered in [14] and also studied in [11]. An overview of these estimators is given in Section 2.2.

The proposed estimators are consistent with the respective estimators when the explanatory variable t take values in Euclidean spaces, that is, in this case, the proposed estimators correspond to the estimators introduced by Engle *et al.* [6].

2.2 *Review of non-parametric estimators on Riemannian manifolds*

2.2.1 *Preliminaries*

Let (M, ξ) be a d-dimensional connected Riemannian manifold. We denote by d_ξ the distance in M induced by ξ. (M, ξ) is a complete Riemannian manifold if (M, d_ξ) is complete as a metric space. $\mathbb{R}^d$ endowed with the Euclidean metric, ξ_{e}^d, the hyperbolic space (H^d, ξ_{h}), the d-dimensional torus (T^d, ξ) and the graph of a smooth function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ are examples of complete Riemannian manifolds.

Let $t \in (M, \xi)$ and T_tM be the tangent fiber of M at t. Let $\exp_t : T_tM \longrightarrow M$ be the exponential map induced by the metric ξ, see for instance [5]. Recall that $\exp_t(0_t) = t$, where 0_t is the null vector of T_tM, and there exists a neighborhood W of 0_t that satisfies that $\exp_t|_W : W \longrightarrow \exp_t(W)$ is a diffeomorphism. We say that a ball $B_r(t) = \{q \in M : d_\xi(t, q) < r\}$ is *normal* if there exists a ball $B_r(0_t)$ such that

$$\exp_t|_{B_r(0_t)} : B_r(0_t) \longrightarrow B_r(p)$$

is a diffeomorphism.

The injectivity radius of (M, ξ) is defined by

$$\mathrm{inj}_\xi(M) := \inf_{r>0}\{B_r(t) \text{ is a normal ball}\}.$$

In the case of the Euclidean or the hyperbolic space, it is easy to see that $inj_{\xi_{\mathrm{e}}^d}(\mathbb{R}^d) = inj_{\xi_{\mathrm{h}}}(H^d) = \infty$. The surface $M := \{x^2 + y^2 - e^{-z} = 0\} \subseteq \mathbb{R}^3$ endowed with the metric induced by ξ_e^3 is a complete Riemannian manifold without a boundary with an injectivity radius equals zero. If (S^d, ξ_0^d) is the d-dimensional sphere with its canonical metric, then $\mathrm{inj}_{\xi_0^d}(S^d) = \pi$. Every compact Riemannian manifold has a positive injectivity radius.

In this paper, we consider only complete Riemannian manifolds without boundary and positive injectivity radius. The assumption about the positive injectivity radius will be clear in Section 3 where we introduce the estimator proposed by Pelletier.

Let $t \in (M, \xi)$ and $\{v_1, \ldots, v_d\}$ be an orthonormal basis of T_tM and let $B_r(t)$ be a normal ball. The exponential map induces a coordinate system (U, ψ) in (M, ξ) as follows. Let $\gamma : T_tM \longrightarrow \mathbb{R}^d$ be defined by

$$\gamma(v) = (z_1, \ldots, z_d) \quad \text{if } v = \sum_{i=1}^{d} z_i v_i,$$

then $U = B_r(t)$ and $\psi : B_r(t) \longrightarrow \mathbb{R}^d$ is given by

$$\psi(s) = \gamma \circ (\exp_t|_{B_r(0_t)})^{-1}(s).$$

The coordinate system (U, ψ) is called a *normal chart centered at* t.

If $s \in U$, let us denote by $\{\partial/\partial\psi_1|_s, \ldots, \partial/\partial\psi_d|_s\}$ the basis of T_sM induced by (U, ψ), that is, let $q \in B_r(0_t)$ such that $\exp_t(q) = s$, then $\partial/\partial\psi_i|_s$ is the velocity at time zero of the curve $\alpha_i(a) = \exp_t(q + a.v_i)$.

For any $s \in M$, the metric ξ restricted to the tangent fiber of M at s is a symmetric positive defined bilinear form $\xi_s : T_sM \times T_sM \longrightarrow \mathbb{R}$. If (U, ξ) is a coordinate system in M and $s \in U$, then let the smooth matricial map $V_{(U,\xi)} : U \longrightarrow \mathbb{R}^{d\times d}$ be given by

$$(V_{(U,\xi)}(s))_{ij} = \xi_s\left(\left.\frac{\partial}{\partial\psi_i}\right|_s, \left.\frac{\partial}{\partial\psi_j}\right|_s\right).$$

Let $t \in M$ and (U, ξ) be a normal chart centered at t. Consider the function $\theta_t : U \longrightarrow \mathbb{R}$:

$$\theta_t(s) = (\det(V_{(U,\xi)}(s)))^{1/2},$$

which is the volume of the parallelepiped $\{\partial/\partial\psi_1|_s, \ldots, \partial/\partial\psi_d|_s\}$. This function is called *the volume density function*. It is not difficult to see that θ_t does not depend on the selection of the normal chart and that $\theta_t(s) = \theta_s(t)$, see [10], for instance.

For the Euclidean space $(\mathbb{R}^d, \xi_e^d)$ and for the cylinder $(S^1 \times \mathbb{R}, \xi_0^1 + \xi_e^1)$, $\theta_t(s) = 1$ for all s and t where the function is well defined. In [10], we calculated the volume density on a sphere, in this case,

$$\theta_s(t) = \frac{|\sin(d_\xi(s,t))|}{d_\xi(s,t)} \quad \text{for } t \neq s, -s \quad \text{and } \theta_s(s) = 1.$$

2.2.2 *The non-parametric estimator*

Let $(y_1, t_1), \ldots, (y_n, t_n)$ be iid random objects that take values on $\mathbb{R} \times M$. In order to estimate $r(\tau) = E(y|t = \tau)$, Pelletier [14] proposed a non-parametric kernel-type estimator. The proposal introduced by Pelletier was built analogous to the kernel-type estimator on (M, ξ) considering the distance d_ξ on M and the volume density function of (M, ξ) in order to take into account the curvature of the manifold. More precisely, the non-parametric estimator can be defined as

$$r_n(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i) y_i \tag{3}$$

with $w_{n,h}(t, t_i) = \theta_t^{-1}(t_i) K(d_\xi(t, t_i)/h) / [\sum_{k=1}^{n} \theta_t^{-1}(t_k) K(d_\xi(t, t_k)/h)]^{-1}$, where $K : \mathbb{R} \to \mathbb{R}$ is a non-negative function, $\theta_t(s)$ is the volume density function on (M, ξ) and the bandwidth h is a sequence of real positive numbers such that $\lim_{n\to\infty} h = 0$ and $h < inj_\xi M$, for all n. The last requirement on the bandwidth guarantees that Equation (3) is well defined for all $t \in M$. Pelletier [14] studied some properties of this estimator such as the asymptotic pointwise mean squared error. On the other hand, Henry and Rodriguez [10] proposed a robust version that generalized this estimator and studied some asymptotic properties.

3. Asymptotic behavior

We begin with the following assumptions required to derive the large sample properties of the proposed estimators. In this section, we study the asymptotic behavior of the regression parameter estimator and the non-parametric component of the model under the following conditions:

(H1) Let M_0 be a compact set on M such that f is a bounded function such that $\inf_{t\in M_0} f(t) = A > 0$ and $\inf_{t,s\in M_0} \theta_t(s) = B > 0$.
(H2) The sequence h is such that $nh^4 \to 0$ and $nh_n^d/\log n \to \infty$ as $n \to \infty$.
(H3) $K : \mathbb{R} \to \mathbb{R}$ is a bounded non-negative Lipschitz function of order 1, with compact support $[0, 1]$ satisfying $\int_{\mathbb{R}^d} K(\|\mathbf{u}\|)\, d\mathbf{u} = 1$, $\int_{\mathbb{R}^d} \mathbf{u} K(\|\mathbf{u}\|)\, d\mathbf{u} = \mathbf{0}$ and $0 < \int_{\mathbb{R}^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|)\, d\mathbf{u} < \infty$.
(H4) For any open set U_0 of M such that $M_0 \subset U_0$, the functions g and ϕ_j for $1 \leq j \leq p$ are of class C^2 on U_0.
(H5) The errors ε_i and η_{ij} for $1 \leq i \leq n$ and $1 \leq j \leq p$ are independent and $E|\varepsilon_1|^r + \sum_{j=1}^{p} E|\eta_{1j}|^r < \infty$ for $r \geq 3$, $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_1) > 0$ and $\Sigma = E(\eta_1^{\mathrm{T}} \eta_1)$ is a positive definite matrix.

We now establish the large sample properties of the estimators. Theorem 3.1 provides the asymptotic normality of $\hat{\boldsymbol{\beta}}$ and the rate of convergence of the non-parametric estimator $\hat{g}(t)$.

THEOREM 3.1 *Under* H1–H5,

(i) $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, \sigma_\varepsilon^2 \Sigma^{-1})$.
(ii) $\sup_{t \in M_0} |\hat{g}(t) - g(t)| = O(h^2) + O(\sqrt{\log n / nh^d})$.

Note that this theorem is consistent with the corresponding results in the Euclidean case.

Remark 3.1 The fact that $\theta_t(t) = 1$ for all $t \in M$ guarantees that the bound of θ in H1 holds. The assumptions H2 and H3 are standard when kernel estimators are considered.

In many practical problems, it is interesting to make inference on the regression parameter, such as the construction of the confidence regions or hypothesis tests. The obtained asymptotic distribution can be used to construct a Wald-type statistic, more precisely, to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. It seems natural to test H_0 through the Wald-type statistic:

$$T_n = \frac{n}{\hat{\sigma}_\varepsilon^2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}} \hat{\Sigma} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

where $\hat{\sigma}_\varepsilon^2$ and $\hat{\Sigma}$ are the estimates of σ_ε^2 and Σ, respectively. For example,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \hat{g}(t_i))^2 \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \hat{\boldsymbol{\phi}}(t))^{\mathrm{T}}(\mathbf{x}_i - \hat{\boldsymbol{\phi}}(t))$$

may be the considered estimators.

Lemma A.2 and Remark A.5 given in Appendix 1 show that $\hat{\Sigma}$ and $\hat{\sigma}_\varepsilon^2$ are consistent with Σ and σ_ε^2, respectively. Therefore, under the null hypothesis, $T_n \xrightarrow{\mathcal{D}} \chi_p^2$. Thus, if we test H_0 at a significance level α, the Wald test rejects H_0 when $T_n > \chi_{p,\alpha}^2$, where $\chi_{p,\alpha}^2$ is the corresponding $1 - \alpha$ quantile of the χ_p^2.

4. Case studies

4.1 *Selection of the smoothing parameter*

An important issue in any smoothing procedure is the choice of the smoothing parameter. Under a non-parametric regression model with carriers in an Euclidean space, that is, when (M, ξ) is $(\mathbb{R}^d, \xi_{\mathrm{e}}^d)$, the two commonly used approaches are L^2 cross-validation and plug-in methods. In this section, we include a cross-validation method for the choice of the bandwidth in the case of partly linear models. The asymptotic properties of data-driven estimators require further careful investigation and are beyond the scope of this paper.

The cross-validation method constructs an asymptotically optimal data-driven bandwidth, and thus adaptive data-driven estimators, by minimizing

$$\mathrm{CV}(h) = \sum_{i=1}^{n}[(y_i - \hat{\phi}_{0,-i,h}(t_i)) - (\mathbf{x}_i - \hat{\boldsymbol{\phi}}_{-i,h}(t_i))^{\mathrm{T}}\tilde{\boldsymbol{\beta}}]^2,$$

where $\hat{\phi}_{0,-i,h}(t)$ and $\hat{\boldsymbol{\phi}}_{-i,h}(t) = (\hat{\phi}_{1,-i,h}(t), \ldots, \hat{\phi}_{p,-i,h}(t))$ denote the non-parametric estimators computed with bandwidth h using all the data expect the ith observation and $\widetilde{\boldsymbol{\beta}}$ minimize $\sum_{i=1}^{n}[(y_i - \hat{\phi}_{0,-i,h}(t_i)) - (\mathbf{x}_i - \hat{\boldsymbol{\phi}}_{-i,h}(t_i))^{\mathrm{T}}\boldsymbol{\beta}]^2$ in $\boldsymbol{\beta}$.

4.2 *Simulation study*

To evaluate the performance of the estimation procedure, we conducted a simulation study. We considered two models in two different Riemannian manifolds, the sphere and the cylinder endowed with the metric induced by the canonical metric of $\mathbb{R}^3$. We performed 1000 replications of independent samples of size $n = 50, 100, 150$ and 200 according to the following models:

Sphere case: The variables (y_i, x_i, t_i) for $1 \leq i \leq n$ were generated as

$$y_i = \beta\, x_i + \exp\{-(t_{i1} + 2t_{i2} + t_{i3})^2\} + \varepsilon_i \quad \text{and} \quad x_i = t_{i1} + t_{i2} + t_{i3} + \eta_i,$$

where $t_i = (\cos(\theta_i)\cos(\gamma_i), \sin(\theta_i)\cos(\gamma_i), \sin(\gamma_i))$ with θ_i and γ_i follow a von Mises distribution with means 0 and π and concentration parameters 3 and 5, respectively. In this case, the functions $g(t)$ and $\hat{\boldsymbol{\phi}}(t)$ are equal to $\exp\{[(1,1,1)^{\mathrm{T}}t]^2\}$ and $(1,1,1)^{\mathrm{T}}t$, respectively.

Cylinder case: The variables (y_i, x_i, t_i) for $1 \leq i \leq n$ were generated as

$$y_i = \beta\, x_i + s_i^2 + \sin(\theta_i) + \varepsilon_i \quad \text{and} \quad x_i = \exp(\theta_i) + \eta_i,$$

where $t_i = (\cos(\theta_i), \sin(\theta_i), s_i)$ with the variables θ_i follow a von Mises distribution with mean π and concentration parameter 3 and the variables s_i are uniform in $(-2, 2)$, that is, t_i have support in the cylinder with radius 1 and height between $(-2, 2)$. Note that in this model, $g(t) = (e_3^{\mathrm{T}}t)^2 + \sin(\arctan(e_2^{\mathrm{T}}t/e_1^{\mathrm{T}}t))$ and $\hat{\boldsymbol{\phi}}(t) = \exp(\arctan(e_2^{\mathrm{T}}t/e_1^{\mathrm{T}}t))$ where e_i for $i = 1, 2, 3$ are the canonical vectors of $\mathbb{R}^3$.

In all cases, the regression parameter $\boldsymbol{\beta}$ was taken to be equal to 5 and the errors ε_i and η_i were iid normal with mean zero and standard deviation 1. In the smoothing procedure, the kernel was taken as $K(u) = 30\, u^2(1-u)^2 I_{(0,1)}(u)$. With respect to the selection of the smoothing parameter, we applied the cross-validation procedure described in Section 4.1. Furthermore, to analyze the effect of the bandwidth on the estimation procedure, we computed the estimators on a grid of bandwidths. We considered an equispaced grid of length 10 between 0.5 and π in the sphere case and between 1 and 2π in the cylinder case. The distance d_ξ for these manifolds can be found in [10,11] and the volume density function in Section 2.2.1. Tables 1 and 2 give the mean, standard deviation (sd), and mean square error (MSE) for the regression estimates of β and the mean of the MSEs of the regression function g over the 1000 replications when we consider the cross-validation procedure. Tables 3 and 4 report the MSE for the regression estimates of β and

Table 1. Performance of $\hat{\boldsymbol{\beta}}$ and $\hat{g}$ in the sphere case using cross-validation.

n	Mean($\hat{\boldsymbol{\beta}}$)	sd($\hat{\boldsymbol{\beta}}$)	MSE($\hat{\boldsymbol{\beta}}$)	MSE($\hat{g}$)
50	5.1103	0.1411	0.0321	0.2102
100	5.1147	0.0993	0.0230	0.1846
150	5.1023	0.0907	0.0187	0.1709
200	5.1024	0.0807	0.0170	0.1644

Table 2. Performance of $\hat{\boldsymbol{\beta}}$ and $\hat{g}$ in the cylinder case using cross-validation.

n	Mean($\hat{\boldsymbol{\beta}}$)	sd($\hat{\boldsymbol{\beta}}$)	MSE($\hat{\boldsymbol{\beta}}$)	MSE($\hat{g}$)
50	4.9836	0.0166	0.0005	0.1756
100	4.9866	0.0119	0.0003	0.1545
150	4.9873	0.0097	0.0003	0.1473
200	4.9877	0.0086	0.0002	0.1397

Table 3. Performance of $\hat{\boldsymbol{\beta}}$ and $\hat{g}$ in the sphere case for different bandwidths.

	Bandwidths									
	0.5	0.793	1.087	1.380	1.674	1.967	2.261	2.554	2.848	3.141
$n = 50$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0311	0.0248	0.0218	0.0207	0.0253	0.0256	0.0293	0.0300	0.0325	0.0337
MSE($\hat{g}$)	0.2715	0.1698	0.1620	0.1772	0.1929	0.2038	0.2102	0.2126	0.2112	0.2130
$n = 100$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0119	0.0116	0.0102	0.0118	0.0146	0.0164	0.0191	0.0209	0.0211	0.0233
MSE($\hat{g}$)	0.1557	0.1134	0.1256	0.1527	0.1747	0.1861	0.1916	0.1936	0.1918	0.1891
$n = 150$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0082	0.0071	0.0069	0.0079	0.0112	0.0143	0.0170	0.0172	0.0192	0.0202
MSE($\hat{g}$)	0.1142	0.0932	0.1150	0.1437	0.1672	0.1792	0.1889	0.1866	0.1863	0.1832
$n = 200$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0056	0.0052	0.0052	0.0072	0.0099	0.0127	0.0155	0.0166	0.0183	0.0196
MSE($\hat{g}$)	0.0931	0.0820	0.1092	0.1388	0.1640	0.1768	0.1835	0.1856	0.1833	0.1818

Table 4. Performance of $\hat{\boldsymbol{\beta}}$ and $\hat{g}$ in the cylinder case for different bandwidths.

	Bandwidth									
	1	1.454	1.908	2.362	2.816	3.270	3.724	4.178	4.632	5.086
$n = 50$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0006	0.0005	0.0005	0.0006	0.0007	0.0007	0.0007	0.0008	0.0008	0.0007
MSE($\hat{g}$)	0.9471	0.5392	0.2428	0.1936	0.2413	0.2786	0.3049	0.3122	0.3158	0.3089
$n = 50$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0003	0.0003	0.0004	0.0004	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
MSE($\hat{g}$)	0.5914	0.4312	0.2269	0.1797	0.2283	0.2622	0.2855	0.2930	0.2935	0.2987
$n = 50$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0002	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
MSE($\hat{g}$)	0.4983	0.3567	0.2214	0.1740	0.2232	0.2592	0.2822	0.2891	0.2903	0.2917
$n = 50$										
MSE($\hat{\boldsymbol{\beta}}$)	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
MSE($\hat{g}$)	0.4476	0.3222	0.2191	0.1713	0.2210	0.2573	0.2786	0.2858	0.2857	0.2882

the mean of the MSEs of the regression function g over the 1000 replications for each bandwidth considered.

From Tables 1 and 2, we can see that the estimators in the two considered schemes show a good behavior. In both the cases, the mean of the MSEs of the parametric and non-parametric estimators is small and reflects the good performance of the proposed estimators. As expected, it is interesting to note that when the sample size increases, the performance of the MSEs of $\boldsymbol{\beta}$ and g is even better. This effect is also summarized in Tables 3 and 4. From these tables, we can see that the behavior of the proposed estimators is stable across the bandwidths considered, except, as expected, when the bandwidths and the sample sizes are small.

4.3 *Application to real data*

In this section, we apply a partly linear model to an environmental data set in order to study the atmospheric SO_2 pollution incidents. The variables included in the study are the direction

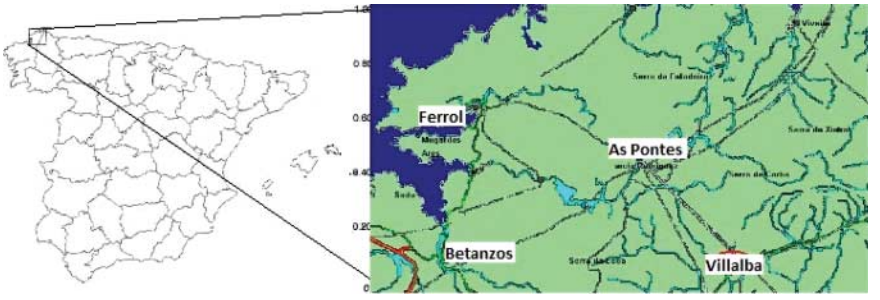

Figure 1. Location of Vilalba, Galicia, Spain.

and speed of the wind, temperature and the SO_2 concentration in the meteorologic station at Vilalba (Lugo in Galicia, Spain). García-Jurado [8] and Prada-Sanchez [16] applied models to the prediction of atmospheric SO_2 pollution incidents in the vicinity of the coal/oil-fired power station at As Pontes, A Coruña, Galicia, Spain. They observed that the prediction of the SO_2 time series is challenging because they consist in near-zero values interrupted between a few days and several weeks by episodes lasting a few hours at random in which values rise to high levels and then fall to zero (Figure 1).

Therefore, in order for the predictions to be based on data representing a reasonably large number of incidents, we reproduced the construction of the historical matrix introduced in [16]. For this purpose, we took as samples 1000 rows of the historical matrix that was constructed and updated as follows.

First, we determined the range of 2-hour means observed during the previous 2 years. Then, we divided the non-near-zero region of this range into 10 strata containing approximately equal numbers of values, randomly selected 100 values y_l from each stratum and associated each with the corresponding predictor values x_l and t_l. The 1000 $(1 + p + d)$-tuplets so formed made up the 1000 rows of the seed of the historical matrix. Thereafter, during on-line processing, the historical matrix was updated whenever a non-near-zero y_k occurred by identifying the stratum to which y_k belonged and substituting x_k and t_k for the oldest row of the historical matrix at time $k - 1$ belonging to this stratum.

The data were recorded daily each minute during the year 2009 and we considered a 1500-row historical matrix. The variables that we considered in the model are given in Table 5.

Note that the variables $t_i = (t_{1i}, t_{2i})$ have support in the cylinder. The maximum of the wind speed in this case was 7.7, so we considered the variable t to belong in a cylinder of height between 0 and 10. Therefore, we modeled the response variable using the following model: $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + g(t_i) + \varepsilon_i$.

In the smoothing procedure, we considered the same kernel that we used in the simulation study and we chose the bandwidth using a cross-validation procedure. Because of the computational burden of the cross-validation method, and because there is really no need of using this method with a sample as large as 1500, we also determined h by the split sample method, that is, by dividing the

Table 5. Environmental variables considered in the model.

y_i	SO_2 emission is measured in $\mu g/m^3$
x_{1i}	SO_2 emission in the instant $i - 30$
x_{2i}	SO_2 emission difference between the instants $i - 35$ and $i - 30$
x_{3i}	The temperature in °C
t_{1i}	Wind direction in radians from the north
t_{2i}	Wind speed in m/s

Table 6. Estimates of the regression parameter.

$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_2$	$\hat{\boldsymbol{\beta}}_3$
0.7988	0.5766	−0.0013

historical matrix into a 750-member training set with an odd index and a 750-member validation set with an even index and taking for h the value minimizing

$$\mathrm{SV}(h) = \sum_{i=1}^{[n/2]} [(y_{2i} - \hat{\phi}_{0,E,h}(t_{2i})) - (\mathbf{x}_{2i} - \hat{\boldsymbol{\phi}}_{E,h}(t_{2i}))^{\mathrm{T}} \tilde{\boldsymbol{\beta}}]^2,$$

where $\hat{\boldsymbol{\phi}}_{E,h}(t) = (\hat{\phi}_{1,E,h}(t), \ldots, \hat{\phi}_{p,E,h}(t))$ and $\hat{\phi}_{0,E,h}(t)$ denote the non-parametric estimators computed with the bandwidth h using the data with the odd index and $\tilde{\boldsymbol{\beta}}$ minimize $\sum_{i=1}^{[n/2]} [(y_{2i} - \hat{\phi}_{0,E,h}(t_{2i})) - (\mathbf{x}_{2i} - \hat{\boldsymbol{\phi}}_{E,h}(t_{2i}))^{\mathrm{T}} \boldsymbol{\beta}]^2$ in $\boldsymbol{\beta}$. In this case, the selected bandwidth was $h_{\mathrm{sv}} = 2.1$. Table 6 reports the values of the estimates of the regression parameters. Figures 2 and 3 show the estimates of the regression function over a grid of 1200 points in the cylinder. The graphics are quite similar. The differences arise only in the form the function has been plotted. Figure 4 shows the residuals versus order in each case of the non-parametric model considered. In particular, Figure 4(a) shows the residuals $x_{1i} - \hat{\phi}_1(t_i)$ against i, Figure 4(b) and (c) the same plots related

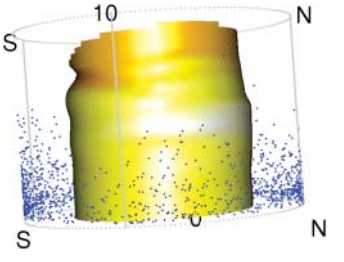

Figure 2. Estimates of the regression function over the cylinder.

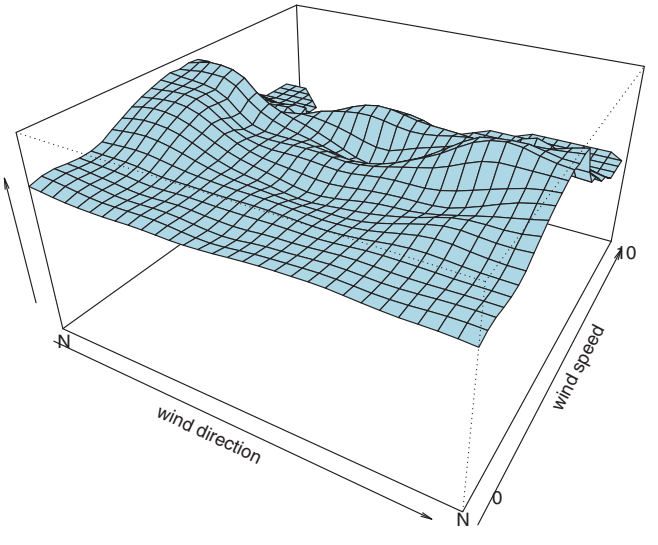

Figure 3. Estimates of the regression function projected in the plane.

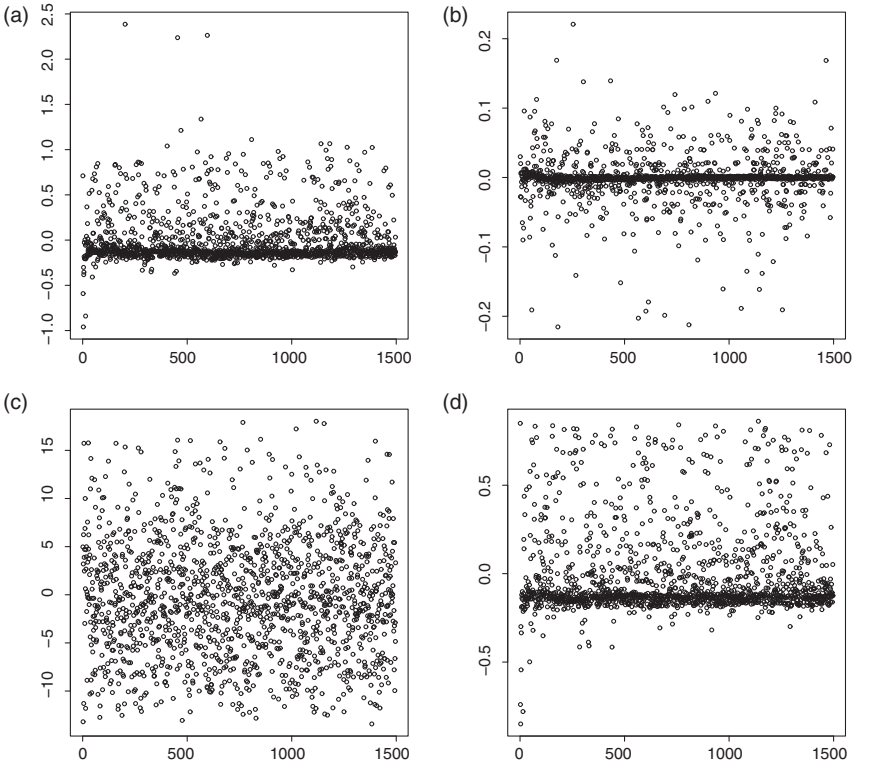

Figure 4. Residuals versus order plots: (a) $x_{1i} - \hat{\phi}_1(t_i)$ versus i, (b) $x_{2i} - \hat{\phi}_2(t_i)$ versus i, (c) $x_{2i} - \hat{\phi}_2(t_i)$ versus i and (d) $y_i - \hat{\phi}_0(t_i)$ versus i.

to x_{2i} and x_{3i}, respectively, and Figure 4(d) $y_i - \hat{\phi}_0(t_i)$ against i. These plots allow us to verify the hypothesis of independence assumed over the errors.

As mentioned in Section 3, we conducted a Wald-type test for the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. If we consider the hypothesis $H_0 : \boldsymbol{\beta} = (0.8, 0.6, 0)$, the test statistic $T_n = 1.59$ and p-value $= 0.6594$. Therefore, we did not reject the null hypothesis. Thus, we can conclude that temperature does not have a significance impact on the pollution incidents.

To evaluate the performance of the partly linear model, we considered a full non-parametric model to explain y_i based on the variables x_{1i}, x_{2i} and x_{3i} through an unknown function η. In this

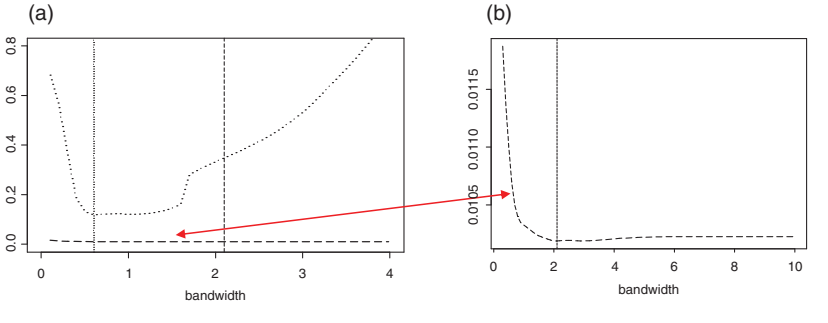

Figure 5. Comparison of the errors: the dotted line corresponds to the full non-parametric model and the dashed line to the partly linear model. The vertical lines correspond to the optimal bandwidths in each model.

case, we used the Nadaraya–Watson estimator with a quadratic kernel. In the smoothing procedure, as the matrix of the bandwidth, we considered a multiple of the identity matrix. We compared the prediction error for both the models computing, in the case of the full non-parametric model, $\mathrm{EP}(h) = \sum_{i=1}^{[n/2]}[(y_{2i} - \hat{\eta}(x_{1,2i}, x_{2,2i}, x_{3,2i}))]^2$ for a grid of 100 equispaced bandwidths between 0.1 and 4. For the partly linear model, we computed $SV(h)$ for the same grid of bandwidth. As can see from Figure 5(a), the partly linear model has a better predictive power than the full non-parametric model. The comparison of the errors in the figure seems to be in a different scale and the proposed method seems to be insensitive with respect to the bandwidth. Figure 5(b) shows only the prediction error of the proposed estimators. Here, we can see that the behavior for small values of bandwidth seems to be more unstable but better than that observed for the full non-parametric model.

Acknowledgements

We thank three anonymous referees and the Editor for valuable comments, which has lead to an improved version of the original paper. We thank María Leyenda Rodríguez for the preparation of the data set. This work was done when the second and third authors were visiting the University of Santiago de Compostela, and they are grateful to all statistics groups for their kind hospitality. This research was partly supported by Grants X-018 from Universidad de Buenos Aires, PIP 1122008010216 from CONICET and PICT -00821 from ANPCYT, Argentina, and also by Spanish Grants MTM2008-03010/MTM of the Ministerio Español de Ciencia e Innovación and XUGA Grants PGIDIT07 PXIB207031PR.

References

[1] G. Aneiros-Pérez and G. Quintela del RÌo, *Plug-in bandwidth choice in partial linear regression models with autoregressive errors*, J. Statist. Plann. Inference 57 (2002), pp. 23–48.
[2] G. Aneiros-Pérez and F. Vieu, *Semi-functional partial linear regression*, Stat. Probab. Lett. 76 (2006), pp. 1102–1110.
[3] R. Bhattacharya and V. Patrangenaru, *Nonparametric estimation of location and dispersion on Riemannian manifolds*, J. Statist. Plann. Inference 108 (2002), pp. 23–35.
[4] H. Chen, *Convergence rates for parametric components in a partly linear model*, Ann. Statist. 16 (1988), pp. 136–146.
[5] M. Do Carmo, *Riemannian Geometry*, Projeto Euclides, IMPA, Brazil, 2005.
[6] R. Engle, C. Granger, J. Rice, and A. Weiss, *Semiparametric estimates of the relation between weather and electricity sales*, J. Amer. Statist. Assoc. 81 (1986), pp. 310–320.
[7] F. Ferraty and F. Vieu, *Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination*, J. Nonparametr. Stat. 16 (2004), pp. 111–125.
[8] I. García-Jurado, W. Gonzalez-Manteiga, J.M. Prada-Sanchez, M. Febrero-Bande, and R. Cao, *Predicting using Box-Jenkins, nonparametric and bootstrap techniques*, Technometrics 37 (1995), pp. 303–310.
[9] H. Hendriks and Z. Landsman, *Asymptotic data analysis on manifolds*, Ann. Statist. 35(1) (2007), pp. 109–131.
[10] G. Henry and D. Rodriguez, *Robust nonparametric regression on Riemannian manifolds*, J. Nonparametr. Stat. 21(5) (2009), pp. 611–628.
[11] G. Henry and D. Rodriguez, *Kernel density estimation on Riemannian manifolds: Asymptotic results*, J. Math. Imaging Vision 43 (2009), pp. 235–639.
[12] H. Liang, *Asymptotic normality of nonparametric part in partially linear models with measurement error in the nonparametric part*, J. statist. plann. inference 86 (2000), pp. 51–62.
[13] K. Mardia, *Statistics of Directional Data*, Academic Press, London, 1972.
[14] B. Pelletier, *Nonparametric regression estimation on closed Riemannian manifolds*, J. Nonparametr. Stat. 18 (2006), pp. 57–67.
[15] X. Pennec, *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*, J. Math. Imaging Vision 25 (2006), pp. 127–154.
[16] J.M. Prada-Sanchez, M. Febrero, T. Cotos-Yañez, W. Gonzalez-Manteiga, J. Bermudez-Cela, and T. Lucas-Dominguez, *Prediction of SO_2 pollution incidents near a power station using partially linear models and an historical matrix of predictor-response vectors*, Environmetrics 11 (2000), pp. 209–225.
[17] P. Speckman, *Kernel smoothing in partial linear models*, J. Roy. Statist. Soc. Ser. B 50 (1988), pp. 413–436.

Appendix 1

LEMMA A.1 *Let* $\tilde{\phi}_j(t) = \phi_j(t) - \sum_{i=1}^n w_{n,h}(t,t_i)x_{ij}$ *for* $1 \le j \le p$ *and* $\tilde{\phi}_0(t) = \phi_0(t) - \sum_{i=1}^n w_{n,h}(t,t_i)y_i$. *Under* H1–H4, *we have that*

$$\sup_{t\in M_0} |\tilde{\gamma}(t)| = O(h^2) + O\left(\sqrt{\frac{\log n}{nh^d}}\right) \quad a.s.,$$

where $\tilde{\gamma} \in \{\tilde{\phi}_j;\ 0 \le j \le p\}$.

Proof of Lemma A.1 Let $\gamma \in \{\phi_j;\ 0 \le j \le p\}$. To fix the idea, we consider $\gamma = \phi_j(t)$ for $1 \le j \le p$, the case $j = 0$ is similar, then

$$\begin{aligned}\tilde{\gamma}(t) &= \phi_j(t) - \sum_{i=1}^n w_{n,h}(t,t_i)x_{ij} \\ &= \frac{(1/nh^d)\sum_{i=1}^n (1/\theta_t(t_i))K(d_\xi(t,t_i)/h)(\phi_j(t) - x_{ij})}{(1/nh^d)\sum_{i=1}^n (1/\theta_t(t_i))K(d_\xi(t,t_i)/h)}.\end{aligned}$$

By H1, $\inf_{t\in M}(1/h^d)E((1/\theta_t(t_1))K(d_\xi(t,t_1)/h)) \ge A > 0$ and the strong uniform consistency of $\hat{f}_n(t) = (nh^d)^{-1}\sum_{k=1}^n \theta_t^{-1}(t_k)\ K(d_\xi(t,t_k)/h)$ obtained in [11], we can concentrate only the numerator of $\tilde{\gamma}(t)$.

Using the results obtained in [11,14], we have that

$$\sup_{t\in M_0}\left|E\left(\frac{1}{nh^d}\sum_{i=1}^n \frac{1}{\theta_t(t_i)}K\left(\frac{d_\xi(t,t_i)}{h}\right)(\phi_j(t) - x_{ij})\right)\right| = O(h^2), \tag{A1}$$

$$\sup_{t\in M_0}\left|\mathrm{var}\left(\frac{1}{nh^d}\sum_{i=1}^n \frac{1}{\theta_t(t_i)}K\left(\frac{d_\xi(t,t_i)}{h}\right)x_{ij}\right)\right| = O(nh^d). \tag{A2}$$

Finally, the proof follows analogous to the proof of Lemma 3.1 reported in [7]. ■

LEMMA A.2 *Under* H1–H4, *we have that* $n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{x}}_i^{\mathrm{T}}\tilde{\boldsymbol{x}}_i \xrightarrow{\mathrm{p}} \Sigma$, *where* $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \hat{\boldsymbol{\phi}}(t_i)$.

Proof of Lemma A.2 The elements l and s of $n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i^{\mathrm{T}}\tilde{\mathbf{x}}_i$ can be written as

$$\left(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i^{\mathrm{T}}\tilde{\mathbf{x}}_i\right)_{ls} = n^{-1}\left(\sum_{i=1}^n \eta_{il}\eta_{is} + \sum_{i=1}^n \tilde{\phi}_l(t_i)\eta_{is} + \sum_{i=1}^n \tilde{\phi}_s(t_i)\eta_{il} + \sum_{i=1}^n \tilde{\phi}_l(t_i)\tilde{\phi}_s(t_i)\right),$$

where $\tilde{\phi}_j(t) = \phi_j(t) - \hat{\phi}_j(t)$. We need to show that all terms except the first term converge to zero. Applying the strong law of large numbers, we get that $n^{-1}\sum_{i=1}^n \eta_{il}\eta_{is} \xrightarrow{\mathrm{p}} \Sigma_{ls}$.

By Lemma A.1, the fact that $n^{-1}\sum_{i=1}^n \eta_{il}^2 \xrightarrow{\mathrm{p}} \Sigma_{ll}$ and using the Cauchy–Schwarz inequality, we get the result. ■

LEMMA A.3 *Under* H1–H3, *we have that* $\sup_{t\in M}\max_{1\le j\le n}|w_{n,h}(t,t_j)| = O((nh^d)^{-1})$.

Proof of Lemma A.3 Note that

$$w_{n,h}(t,t_j) = \frac{(1/nh^d)(1/\theta_t(t_j))K(d_\xi(t,t_j)/h)[(1/h^d)E((1/\theta_t(t_1))K(d_\xi(t,t_1)/h))]^{-1}}{(1/nh^d)\sum_{i=1}^n (1/\theta_t(t_i))K(d_\xi(t,t_i)/h)[(1/h^d)E((1/\theta_t(t_1))K(d_\xi(t,t_1)/h))]^{-1}}.$$

According to Henry and Rodriguez [11], we have that

$$\sup_{t\in M}\left|\frac{1}{nh^d}\sum_{i=1}^{n}\frac{1}{\theta_t(t_i)}K\left(\frac{d_\xi(t,t_i)}{h}\right)\left[\frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K\left(\frac{d_\xi(t,t_1)}{h}\right)\right)\right]^{-1}-1\right|=o(1)\quad\text{a.s.,}\tag{A3}$$

$$\inf_{t\in M}\frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K\left(\frac{d_\xi(t,t_1)}{h}\right)\right)\geq A>0.\tag{A4}$$

Then, by Equations (A3) and (A4) and the boundedness of K and θ_t, the lemma holds. ■

Remark A.4 Note that by Lemmas A.1 and A.3 and using Lemma A.1 reported in [12], we have that

$$\max_{1\leq i\leq n}\left|\gamma(t_i)-\sum_{k=1}^{n}w_{n,h}(t_i,t_k)\gamma(t_k)\right|=O(h^2)+O\left(\sqrt{\frac{\log n}{nh^d}}\right)\quad\text{a.s.}$$

for any $\gamma\in\{\phi_j;\ 0\leq j\leq p\}$.

Proof of Theorem 3.1 (i) We can write $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})=(n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{x}}_i^{\mathrm{T}}\tilde{\mathbf{x}}_i)^{-1}n^{-1/2}[A_{1n}-A_{2n}+A_{3n}]$, where

$$A_{1n}=\sum_{i=1}^{n}\tilde{\mathbf{x}}_i g^*(t_i),\quad A_{2n}=\sum_{i=1}^{n}\tilde{\mathbf{x}}_i\left(\sum_{i=1}^{n}w_{n,h}(t_i,t_j)\varepsilon_j\right),\quad A_{3n}=\sum_{i=1}^{n}\tilde{\mathbf{x}}_i\varepsilon_i,$$

and $g^*(t)=g(t)-\sum_{i=1}^{n}w_{n,h}(t,t_i)g(t_i)$. Using Lemmas A.1–A.3, the asymptotic behavior of A_{1n}, A_{2n} and A_{3n} can be obtained in the same way as in [2].

Specifically, considering the assumptions imposed on h, we can obtain

$$\begin{aligned}A_{1n}&=O(nh^4+h^{-d}\log^2 n)+O(n^{1/2}h^2\log n+h^{-d/2}\log^2 n)+O(n^{1/2}h^2h^{-d/2}\log n)\\&\quad+O(h^{-d}\log^2 n)=o(n^{1/2}),\\A_{2n}&=O(n^{1/2}h^2h^{-d/2}\log n+h^{-d}\log^2 n)+O(h^{-d/2}\log^2 n)+O(h^{-d}\log^2 n)\\&=o(n^{1/2})\end{aligned}$$

and

$$\begin{aligned}A_{3n}&=O(n^{1/2}h^2\log n)+O(h^{-d/2}\log^2 n)+\sum_{i=1}^{n}\eta_i\varepsilon_i+O(h^{-d/2}\log^2 n)\\&=\sum_{i=1}^{n}\eta_i\varepsilon_i+o(n^{1/2}).\end{aligned}$$

Finally, the central limit theorem gives the desired result.

(ii) Note that $\hat{g}(t)-g(t)=\hat{\phi}_0(t)-\phi_0(t)+(\hat{\boldsymbol{\phi}}(t)-\hat{\boldsymbol{\phi}}(t))^{\mathrm{T}}\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{\phi}}(t)^{\mathrm{T}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$. Therefore, Lemma A.1 and part (i) of this theorem allow us to complete the proof. ■

Remark A.5 Finally, we note that Lemma A.1 and Theorem 3.1 state the consistency of the estimator $\hat{\sigma}_\varepsilon^2$ of σ_ε^2 defined in Section 3. More precisely, note that

$$\hat{\sigma}_\varepsilon^2=\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i+\mathbf{x}_i^{\mathrm{T}}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})+g(t_i)-\hat{g}(t_i))^2.$$

Thus, if we distribute the terms, all of them converge to zero by Lemma A.1 and Theorem 3.1 except $(1/n)\sum_{i=1}^{n}\varepsilon_i^2$, since by the strong law of large numbers, it converges to σ_ε^2.