Partially linear models on Riemannian manifolds

Wenceslao Gonzalez-Manteiga[1], Guillermo Henry[2] and Daniela Rodriguez[2]
[1] *Universidad de Santiago de Compostela, Spain*
[2] *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*

Abstract

In partially 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 $(\mathbf{x}^{\mathrm{T}}, t)$. In this paper, estimators of $\boldsymbol{\beta}$ and g are constructed when the explanatory variables t take values on a Riemannian manifold. Our proposal combine the flexibility of these models with the complex structure of a set of explanatory variables. We prove that the resulting estimator of $\boldsymbol{\beta}$ is asymptotically normal under the suitable conditions. Through a simulation study, we explored the performance of the estimators. Finally, we applied the studied model to an example based on real dataset.

Key words and phrases: Nonparametric estimation, Partly linear models, Riemannian manifolds.

1 Introduction

The partially linear models was introduced by [6] to analyzed the relationship between the electricity usage and average daily temperature. In recent years, this model has gained a lot of attention in order to explore the nature of complex nonlinear phenomena. This model has been widely studied in the literature see for example [16], [5], [1] among others. The partially linear models allow modeling the response variable with a set of predictors that enter linearly in the model while one of them is considered in the model nonparametrically.

However, in many applications, the predictors variables take values on a Riemannian manifold more than on Euclidean space and this structure of the variables needs to be taken into account in the estimation procedure. Some examples could be found in meteorology, astronomy, geology and other fields, that include distributions on spheres, tangent bundles, Lie groups, etc. Research on the statistical analysis of variables with some one of this structures was studied by [4], [12] and more recently by [8], [14],[13] and [9].

The aim of this work is to study the partially linear models when the explanatory variable t takes values on a Riemannian manifold, i.e. when the variable to be modeled in a nonparametric way is in a manifold. Our proposal combine the flexibility for these models with the complex structure of a set of explanatory variables.

This paper is organized as follows. In Section 2, we construct estimates for this models and give a brief summary of the nonparametric estimation on Riemannian manifolds proposed in [13]. In Section 3, we present the asymptotic distribution of the regression parameter under regular assumptions on the bandwidth sequence. In Section 4, we explored the performance of the estimators with a simulation study and we show an example using real data. Also, we review a cross validation procedure for partial linear models. Proofs are given in the Appendix.

2 Estimators

2.1 Model and estimators

Let $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)$ be an i.i.d. random vectors valued in $I\!R^{p+1} \times M$ with identically distribution to $(y, \mathbf{x}^{\mathrm{T}}, t)$, where (M, g) is a Riemannian manifolds of dimension d. The partially linear model assume that the relation between the response variable y_i and the covariates $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ can be represented as

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

where the errors ε_i are independent and independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}$, also $E(\varepsilon_i|\mathbf{x}_i, t_i) = 0$. In many situations, it seems reasonable to suppose that a relationship between the covariates $\mathbf{x}$ and t exists, so as in [16] and [1], we will assume that for $1 \le j \le p$

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

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, $y - \phi_0(t) = (\mathbf{x} - \boldsymbol{\phi}(t))^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$. This equation suggest estimate the unknown functions and parameters as follows. Let $\widehat{\phi}_j(t)$ be the nonparametric estimators of ϕ_j for $0 \le j \le p$. Note that the regression functions correspond to predictors taking values in a Riemannian manifold, nonparametric kernel type estimators adapted to this structure was considered in [13] and also studied in [10]. An overview of this estimators can be found in the following Subsection.

Returned to the estimation of the parameter $\boldsymbol{\beta}$, note that using the nonparametric estimators of the functions ϕ_j, the regression parameter can be estimate considering the least square estimators obtained minimizing

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

where $\widehat{\boldsymbol{\phi}}(t) = (\widehat{\phi}_1(t), \ldots, \widehat{\phi}_p(t))$. Then the function g can be estimated as $\widehat{g}(t) = \widehat{\phi}_0(t) - \widehat{\boldsymbol{\phi}}(t)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$. This procedure is consistent with the respective estimators when the explicative variable t take values on Euclidean spaces, i.e. the proposed estimators reduce to know estimators introduced by [6].

2.2 Review of Nonparametric estimators on Riemannian manifolds

2.2.1 Preliminaries

As in [9] we consider (M, g) a $d-$dimensional oriented Riemannian manifold without boundary, complete and with positive injectivity radius ($inj_g M > 0$). From now on, d_g will denote the distance function induced by the metric g. Throughout this note, we will consider the concept of volume density function. For a rigorous definition of this function see [3] or [10]. If we consider the exponential normal chart (U, ψ) of (M, g) induced by an orthonormal basis $\{v_1, \ldots, v_d\}$ of T_sM, then $\theta_s(t) = \left|\det g_t\left(\partial/\partial\psi_i\big|_t, \partial/\partial\psi_j\big|_t\right)\right|^{\frac{1}{2}}$, where $\partial/\partial\psi_i|_t = D_{\alpha_i(0)}exp_s(\dot{\alpha}_i(0))$ with $\alpha_i(u) = exp_s^{-1}(t) + uv_i$ for $t \in U$. For example, when M is $I\!R^d$ with the canonical metric, then $\theta_s(t) = 1$ for all $s, t \in I\!R^d$ and also in the case of the cylinder $\theta_s(t) = 1$. In [9], we calculate the volume density on the sphere, in this case, $\theta_s(t) = |sen(d_g(s, t))|/d_g(s, t)$ for $t \neq s, -s$. and $\theta_s(\pm s) = 1$. See also, [9] for a discussion on the geometric definitions.

2.2.2 The nonparametric estimators

Let $(y_1, t_1), \cdots, (y_n, t_n)$ be i.i.d random objects that take values on $I\!R \times M$. In order to estimate $r(\tau) = E(y|t = \tau)$, Pelletier [13] proposed a nonparametric kernel type estimators. The Pelletiers idea was to build an analogue of a kernel on (M, g), by using a positive function of d_g distance normalized by the volume density function of (M, g), to take into account the curvature of the manifolds. More precisely, the nonparametric estimator can be defined as,

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

with $w_{n,h}(t, t_i) = \theta_t^{-1}(t_i)K(d_g(t, t_i)/h)/[\sum_{k=1}^{n} \theta_t^{-1}(t_k)K(d_g(t, t_k)/h)]^{-1}$ where $K : I\!R \to I\!R$ is a non-negative function, $\theta_t(s)$ the volume density function on (M, g) and the bandwidth h is a sequence of real positive numbers such that $\lim_{n\to\infty} h = 0$ and $h < inj_g M$, for all n. This last requirement on the bandwidth guarantees that (3) is defined for all $t \in M$. In [13], is derived an expression for the asymptotic pointwise bias and variance as well as an expression for the asymptotic integrated mean square error. On the other hand, in [9] is proposed a robust version that generalized these estimators and it is obtained the uniform almost sure consistency over compact set and derived the asymptotic distribution.

3 Asymptotic behavior

The theorem of this section studies the asymptotic behavior of the regression parameter estimator 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 $n \to \infty$.

$H3$. $K : I\!R \to I\!R$ is a bounded nonnegative Lipschitz function of order one, with compact support $[0,1]$ satisfying: $\int_{I\!R^d} K(\|\mathbf{u}\|)d\mathbf{u} = 1$, $\int_{I\!R^d} \mathbf{u}K(\|\mathbf{u}\|)d\mathbf{u} = \mathbf{0}$ and $0 < \int_{I\!R^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|)d\mathbf{u} < \infty$.

$H4$. For any open set U_0 of M_0 such that $M_0 \subset U_0$, the functions g, ϕ_j for $1 \le j \le p$ are of class C^2 on U_0.

$H5$. The errors ε_i and η_{ij} for $1 \le i \le n$ and $1 \le j \le p$ are independent and $E|\varepsilon_1|^r + \sum_{j=1}^p E|\eta_{1j}|^r < \infty$ for $r \ge 3$, $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_1) > 0$ and $\Sigma = E(\eta_1^{\mathrm{T}}\eta_1)$ is a positive defined matrix.

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

Theorem 3.1. Under $H1$ to $H5$ we have that $\sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta) \xrightarrow{\mathcal{D}} N(0, \sigma_\varepsilon^2\Sigma^{-1})$.

Remark 3.2. Note that this theorem is consistent with the respective results in the Euclidean case. The obtained asymptotic distribution can be used to construct a Wald-type statistics to make inference on the regression parameter, that is, when we want to test $H_0 : \boldsymbol\beta = \boldsymbol\beta_0$.

4 Real example and Monte Carlo study

4.1 Selection of the smoothing parameter

An important issue in any smoothing procedure is the choice of the smoothing parameter. Under a nonparametric regression model with carriers in an Euclidean space, i.e., when M is $I\!R^d$ with the canonical metric, two commonly used approaches are L^2 cross–validation and plug–in methods. In this section, we included a cross-validation method for the choice of the bandwidth in the case of partially 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 $CV(h) = \sum_{i=1}^n [(y_i - \widehat{\phi}_{0,-i,h}(t_i)) - (\mathbf{x}_i - \widehat{\boldsymbol\phi}_{-i,h}(t_i))^{\mathrm{T}}\widetilde{\boldsymbol\beta}]^2$, where $\widehat{\phi}_{0,-i,h}(t)$ and $\widehat{\boldsymbol\phi}_{-i,h}(t) = (\widehat{\phi}_{1,-i,h}(t), \ldots, \widehat{\phi}_{p,-i,h}(t))$ denote the nonparametric estimators computed with bandwidth h using all the data expect the $i-$th observation and $\widetilde{\boldsymbol\beta}$ minimize $\sum_{i=1}^n [(y_i - \widehat{\phi}_{0,-i,h}(t_i)) - (\mathbf{x}_i - \widehat{\boldsymbol\phi}_{-i,h}(t_i))^{\mathrm{T}}\boldsymbol\beta]^2$ in $\boldsymbol\beta$.

4.2 Simulation study

To evaluate the performance of the estimation procedure, we conduct a simulation study. We consider two models in two different Riemannian manifolds, the sphere and the cylinder endowed with the metric induced by the canonical metric of $I\!R^3$. We performed 1000 replications of independent samples of size $n = 200$ 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.

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 $(-2, 2)$, i.e. t_i have support in the cylinder with radius 1 and height between $(-2, 2)$.

In all cases, the regression parameter $\boldsymbol{\beta}$ was taken equal 5 and the errors ε_i and η_i are i.i.d. normal with mean 0 and standard deviation 1. In the smoothing procedure, the kernel was taken as the quadratic kernel $K(t) = (15/16)(1 - t^2)^2 I(|x| < 1)$ and we choose the bandwidth using a cross validation procedure described in Section 4.1. The distance d_g for these manifolds can be found in [10] and [9] and the volume density function in Section 2.2.1. Table 4.2.1 give the mean, standard deviations, mean square error for the regression estimates of β and the mean of the mean square error of the regression function g over the 1000 replications.

	mean($\widehat{\boldsymbol{\beta}}$)	sd($\widehat{\boldsymbol{\beta}}$)	MSE($\widehat{\boldsymbol{\beta}}$)	MSE($\widehat{g}$)
sphere case	5.0243	0.0762	0.0064	0.081
cylinder case	4.9845	0.0078	0.0003	0.1001

Table 4.2.1: Performance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{g}$ for both models.

In Table 4.2.1 we can see a good behavior of the estimators in the two considered schemes. In all cases, the mean of the mean square error of the parametric and nonparametric estimators are small and reflect a good performance of the proposed estimators.

4.3 Application to real data

In this Subsection, we applied a partially linear model to an enviroment dataset in order to study the atmospheric SO_2 pollution incidents. The variables included in the study are the direction and the speed of the wind, the temperature and the SO_2 concentration in the meteorologic station at Villalba (Lugo in Galicia, Spain). The data was recorded daily in each minute during the year 2009. The complete dataset has a structure of dependence in the time. Therefore to avoid this dependence we was considered a 2000–row historical matrix that was constructed as in [15]. In a previous work [15] applied a partial linear 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 in Galicia, Spain). But in this case they did not consider the direction of the wind as a directional variable. The variables that we considered in the model was

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 diference between the instant $i - 35$ 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 4.3.1: Enviromente variables considered in the model.

Note that the variables $t_i = (t_{1i}, t_{2i})$ have support in the cylinder. The maximum of the wind speed in this cases is 7.7 then we consider that the variable t belongs in the cylinder of high 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 quadratic kernel and we choose the bandwidth using a cross validation procedure. Because of the computational burden of the cross-validation method, and because there is really no need to use this method with a sample as large as 2000, we also determined h by the split sample method, i.e. by dividing the historical matrix into a 1000-member training set with odd index and a 1000-member validation set with even index, and taking for h the value minimizing

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

where $\widehat{\boldsymbol{\phi}}_{E,h}(t) = (\widehat{\phi}_{1,E,h}(t), \ldots, \widehat{\phi}_{p,E,h}(t))$ and $\widehat{\phi}_{0,E,h}(t)$ denote the nonparametric estimators computed with bandwidth h using the data with even index and $\widetilde{\boldsymbol{\beta}}$ minimize $\sum_{i=1}^{[n/2]} [(y_{2i} - \widehat{\phi}_{0,E,h}(t_{2i})) - (\mathbf{x}_{2i} - \widehat{\boldsymbol{\phi}}_{E,h}(t_{2i}))^{\mathrm{T}} \boldsymbol{\beta}]^2$ in $\boldsymbol{\beta}$. In this case the selected bandwidth was $h_{sv} = 2.5$. Table 4.3.2 reports the estimates values of the regression parameters and the mean and standard deviation of nonparametric estimator $\widehat{g}$ of g. Figure 4.3.1.a) shows the estimate of the regression function over a grid of 1200 points in the cylinder. To evaluate the performance of the partial linear model, we consider a nonparametric model to explain y_i based only in the variables x_{1i} and x_{2i} trough an unknown function η . In this case we estimate with the Naradaya-Watson estimator with quadratic kernel. We compare the prediction error for both models computing, in the case of the full nonparametric model, we compute $EP(h) = \sum_{i=1}^{[n/2]} [(y_{2i} - \widehat{\eta}(x_{1,2i}, x_{2,2i}))]^2$ for a grid of 100 equispaces bandwidth between 0.1 and 10. For the partial linear model we compute the $SV(h)$ for the same grid of bandwidth. As we can see in Figure 4.3.1. b), the partial linear model has a better level predictive and is more stable trough the bandwidth than the full nonparametric model.

$\widehat{\boldsymbol{\beta}}_1$	$\widehat{\boldsymbol{\beta}}_2$	$\widehat{\boldsymbol{\beta}}_3$	Mean($\widehat{g}$)	SD($\widehat{g}$)
0.9728	1.090	-0.0013	0.1141	0.0145

Table 4.3.2: Estimates of regression parameter.

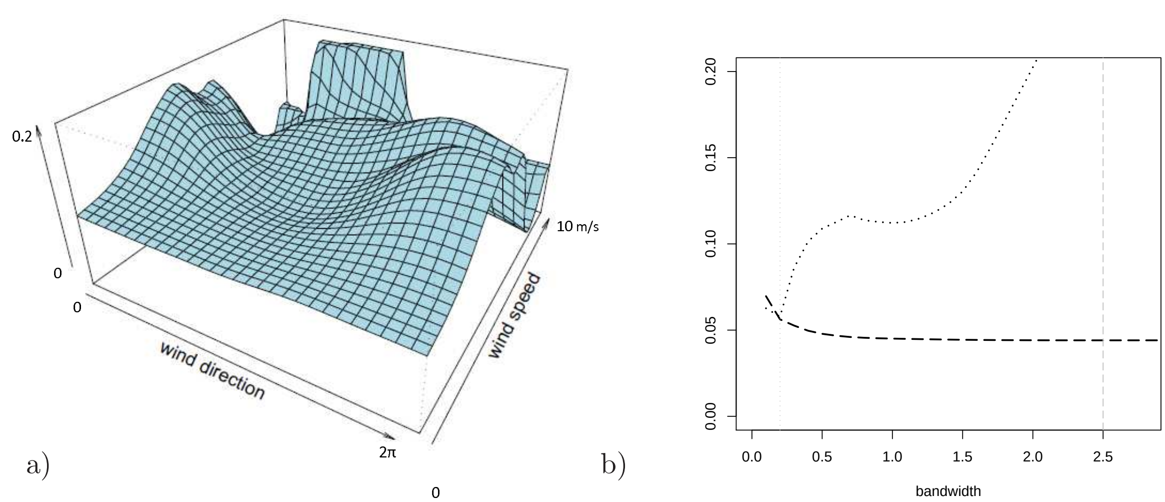

Figure 4.3.1: a) Estimates of the regression functions. b) Comparative of the errors: the dotted line corresponds to the full nonparametric model and the dashed line to the partially linear model. The vertical lines corresponds to the optimal bandwidth in each cases.

Acknowledgments

We would like to thank María Leyenda Rodríguez for the preparation of the dataset. This work was done when the second and third authors were visiting the University of Santiago the Compostela, they are very grateful to all statistics group for their kind hospitality. This research was partially supported by Grants X-018 from the 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 PGIDIT07PXIB207031PR.

A Appendix

Lemma A.1: *Let $\widetilde{\phi}_j(t) = \phi_j(t) - \sum_{i=1}^{b} w_{n,h}(t,t_i)x_{ij}$ for $1 \le j \le p$ and $\widetilde{\phi}_0(t) = \phi_0(t) - \sum_{i=1}^{b} w_{n,h}(t,t_i)y_i$. Under $H1$ to $H4$ we have that $\max_{1\le i \le n} |\widetilde{\gamma}(t_i)| = O(h^2) + O\left(\sqrt{\log n / nh^d}\right)$ a.s. where $\widetilde{\gamma} \in \{\widetilde{\phi}_j; \quad 0 \le j \le p\}$.*

Proof of Lemma A.1: Let $\gamma \in \{\phi_j; \quad 0 \le j \le p\}$ and denote by $\widehat{\gamma}$ the corresponding nonparametric estimator, the $\widetilde{\gamma}(t) = \gamma(t) - \widehat{\gamma}(t)$. Using analogous arguments that those considered in [10] we have that, $\sup_{t\in M_0} |E(\widetilde{\gamma}(t))| = O(h^2)$. Let $s_n = n^2h^{2d} \sup_{t\in M_0} \left|\text{var}(\widehat{\gamma}(t)\widehat{f}_n(t))\right|$ with $\widehat{f}_n(t) = (nh^d)^{-1} \sum_{k=1}^{n} \theta_t^{-1}(t_k)K(d_g(t,t_k)/h)$, by results obtained in [9] we have that $s_n = O(nh^d)$. By $H1$, $\inf_{t\in M} \frac{1}{h^d} E\left(\frac{1}{\theta_t(t_1)}K(d_g(t,t_1)/h)\right) \ge A > 0$. Then, it follows in analogous way that the proof of Lemma 3.1 in [7].□

Lemma A.2: *Under $H1$ to $H4$ we have that $n^{-1}\widetilde{\mathbf{x}}^{\mathrm{T}}\widetilde{\mathbf{x}} \xrightarrow{p} \Sigma$.*

Proof of Lemma A.2: The element l, s of $n^{-1}\widetilde{\mathbf{x}}'\widetilde{\mathbf{x}}$ can be written as

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

where $\widetilde{\phi}_j(t) = \phi_j(t) - \widehat{\phi}_j(t)$. We need to show that all terms except the first term converge to zero and by applying the strong law of large numbers we get that $n^{-1}\sum_{i=1}^{n} \eta_{il}\eta_{is} \xrightarrow{p} \Sigma_{ls}$. Since Lemma A.1 and the fact that $n^{-1}\sum_{i=1}^{n} \eta_{il}^2 \xrightarrow{p} \Sigma_{ll}$ and using the Cauchy-Schwarz inequality we get the result.□

Lemma A.3: *Under $H1$ to $H3$, we have that $\max_{1\le i,j\le n} |w_{n,h}(t_i,t_j)| = O((nh^d)^{-1})$.*

Proof of Lemma A.3: Using the results obtained in [10] and [9] we have that

$$\sup_{t\in M}\left|\frac{1}{nh^d}\sum_{i=1}^{n}\frac{1}{\theta_t(t_i)}K(d_g(t,t_i)/h) - \frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K(d_g(t,t_1)/h)\right)\right| = o(1) \quad \text{a.s.} \tag{4}$$

$$\inf_{t\in M}\frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K(d_g(t,t_1)/h)\right) \ge A > 0. \tag{5}$$

Then by (4) and (5) 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 in [11]; we have that $\max_{1\le i\le n} |\gamma(t_i) - \sum_{k=1}^{n} w_{n,h}(t_i,t_k)\gamma(t_k)| = O(h^2) + O\left(\sqrt{\log n/nh^d}\right)$ a.s. for any $\gamma \in \{\phi_j; \quad 0 \le j \le p\}$.

Proof 3.1: We can write $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (n^{-1}\widetilde{\mathbf{x}}^{\mathrm{T}}\widetilde{\mathbf{x}})^{-1}n^{-1/2}\left[A_{1n} - A_{2n} + A_{3n}\right]$ where

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

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

$$\begin{aligned}
A_{n1} &= O(nh^4 + h^{-d}\log^2 n) + O(n^{1/2}h^2\log n + h^{-d/2}\log^2 n) + O(n^{1/2}h^2h^{-d/2}\log n) + O(h^{-d}\log^2 n) = o(n^{1/2}) \\
A_{n2} &= O(n^{1/2}h^2h^{-d/2}\log n + h^{-d}\log^2 n) + O(h^{-d/2}\log^2 n) + O(h^{-d}\log^2 n) = o(n^{1/2}) \\
A_{n3} &= O(n^{1/2}h^2\log n) + O(h^{-d/2}\log^2 n) + \sum_{i=1}^{n}\eta_i\varepsilon_i + O(h^{-d/2}\log^2 n) = \sum_{i=1}^{n}\eta_i\varepsilon_i + o(n^{1/2})
\end{aligned}$$

Finally, the central limit theorem gives the desired result. □

References

[1] Aneiros-Pérez, G. and Quintela del Río, G. (2002). Plug-in bandwidth choice in partial linear regression models with autoregressive errors. *J. Statist. Planning and Inference.* **57**, 23-48.

[2] Aneiros-Pérez, G. and Vieu, F. (2006). Semi-Functional partial linear regression. *Statistics & Probability Letters.* **76**, 1102-1110.

[3] Besse, A. (1978). Manifolds all of whose Geodesics are Closed. *Springer-Verlag.*

[4] Bhattacharya, R. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference.* **108,** 23-35.

[5] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16,** 136-146.

[6] Engle, R.; Granger, C.; Rice, J. and Weiss, A. (1986). Nonparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81,** 310-320.

[7] Ferraty, F. and Vieu, F. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of Nonparametric Statistics.* **16,** 111-125.

[8] Hendriks, H. and Landsman, Z. (2007). Asymptotic data analysis on manifolds. *Ann. Statist.* **35, 1,** 109-131.

[9] Henry, G. and Rodriguez, D. (2009). Robust Nonparametric Regression on Riemannian Manifolds. *Journal of Nonparametric Statistics.* **21, 5,**611-628.

[10] Henry, G. and Rodriguez, D. (2009). Kernel Density Estimation on Riemannian Manifolds: Asymptotic Results. *Journal Math. Imaging Vis.* **43,** 235-639.

[11] Liang, H. (2000). Asymptotic of nonparametric part in partially linear models with measurement error in the nonparametric part. *Journal of statistical planning and inference.* **86,** 51-62.

[12] Mardia, K. (1972). Statistics of Directional Data. *Academic Press, London.*

[13] Pelletier, B. (2006). Nonparametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics.* **18,** 57-67.

[14] Pennec, X. (2006). Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal Math. Imaging Vis.* **25,** 127-154.

[15] Prada-Sanchez, J.M., Febrero-Bande, M., Cotos-Yañez, T.; Gonzalez-Manteiga, W.; Bermudez-Cela, J. and Lucas-Dominguez, T. (2000). Prediction of SO_2 pollution incidents near a power station using partially linear models and an historical matrix of predictor-response vectors. *Environmetrics.* **11,** 209-225.

[16] Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc.* Ser. B. **50**, 413-436.