Robust estimators in partly linear regression models on Riemannian manifolds

Guillermo Henry and Daniela Rodriguez

October 16, 2014

Abstract

Under a partly linear model we study a family of robust estimates for the regression parameter and the regression function when some of the predictors take values on a Riemannian manifold. We obtain the consistency and the asymptotic normality of the proposed estimators. Simulations and an application to a real dataset show the good performance of our proposal under small samples and contamination.

Non parametric estimation, Partly linear models, Riemannian manifolds, Robustness.

1 Introduction

Partly linear regression models (PLM) assume that the regression function can be modeled linearly on some covariates, while it depends non parametrically on some others. To be more precise, assume that we have a response $y_i \in I\!\!R$ and covariates $(\mathbf{x}_i, t_i)$ such that $\mathbf{x}_i \in I\!\!R^p, t_i \in [0,1]$ satisfying

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

where the errors ε_i are independent and independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$. Since the introductory work by **?**, partly linear models have become an important tool in the modeling of econometric or biometric data, combining the flexibility of non parametric models and the simple interpretations of the linear ones. However, in many applications, some of 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.

2 head one

In a recent paper **?**, PLM are studied when the explanatory variables corresponding to the non parametric component, take values on a Riemannian manifold and the potential application of this model in an environmetric problem is explored. Unfortunately, as shown in Section 5,

this approach may not work well in presence of a small proportion of observations that deviate from the assumed model. One way to avoid this problem is to consider robust estimators that can resist the effect of a small number of atypical observations. The goal of this paper is to introduce resistant estimators for the regression parameter and the regression function under PLM (1), when the predictor variable t takes values on a Riemannian manifold.

This paper is organized as follows. Section 3 gives a brief summary of the classical estimators for this model and introduces the robust estimates. In Section 4, we study the consistency and the asymptotic distribution of the regression parameter under regular assumptions on the bandwidth sequence. Section 5 includes the results of a simulation study conducted in order to explore the performance of the new estimators under normality and contamination. Also, a robust cross validation procedure to select the smoothing parameter is considered. The advantages of the proposed method are also illustrated over a real data set, in Section 6. Proofs are given in the Appendix.

3 The model and the estimators

Assume that we have a sample of n independent variables $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)$ in $I\!\!R^{p+1} \times M$ with the same distribution as $(y, \mathbf{x}^{\mathrm{T}}, t)$, where (M, γ) is a Riemannian manifold of dimension d. As in **?** we consider (M, γ) a $d-$dimensional compact oriented and connected Riemannian manifold without boundary. Note that in this case the injectivity radius of (M, γ) $(inj_\gamma M$) is positive. 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

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + g(t_i) + \varepsilon_i \qquad 1 \le i \le n , \tag{2}$$

where the errors ε_i are independent and independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$. Furthermore, we will assume that ε has symmetric distribution.

3.1 Classical estimators

Denote $\phi_0(\tau) = E(y|t = \tau)$ and $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_p(t))$ where $\phi_j(\tau) = E(x_{ij}|t = \tau)$ for $1 \le j \le p$, then we have that $g(t) = \phi_0(t) - \boldsymbol{\phi}(t)^{\mathrm{T}} \boldsymbol{\beta}$ and hence, $y - \phi_0(t) = (\mathbf{x} - \boldsymbol{\phi}(t))^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon$. The classical least square estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_{ls}$ can be obtained minimizing

$$\widehat{\boldsymbol{\beta}}_{ls} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} [(y_i - \widehat{\phi}_{0,ls}(t_i)) - (\mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{ls}(t_i))^{\mathrm{T}} \boldsymbol{\beta}]^2,$$

where $\widehat{\phi}_{0,ls}$ and $\widehat{\boldsymbol{\phi}}_{ls}$ are non parametric kernel estimators of ϕ_0 and $\boldsymbol{\phi}$, respectively. More precisely, the non parametric estimators $\widehat{\phi}_{0,ls}$ and $\widehat{\phi}_{j,ls}$ of ϕ_0 and ϕ_j can be defined as (see **?**),

$$\widehat{\phi}_{0,ls}(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i) y_i \quad \text{and} \quad \widehat{\phi}_{j,ls}(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i) x_{ij} \tag{3}$$

where $w_{n,h}(t, t_i) = \theta_t^{-1}(t_i)K(d_\gamma(t, t_i)/h_n)/[\sum_{k=1}^n \theta_t^{-1}(t_k)K(d_\gamma(t, t_k)/h_n)]^{-1}$ with $K : I\!R \rightarrow I\!R$ a non-negative function, d_γ the distance induced by the metric γ and $\theta_t(s)$ the volume density function on (M, γ). The bandwidth h_n is a sequence of real positive numbers such that $\lim_{n \to \infty} h_n = 0$ and h_n are smaller than $inj_\gamma M$. For a rigorous definition of the volume density function and the injectivity radius see (Besse, 1978) or **?**.

The final least square estimator of g can be taken as $\widehat{g}_{ls}(t) = \widehat{\phi}_{0,ls}(t) - \widehat{\phi}_{ls}(t)^{\mathrm{T}}\widehat{\beta}_{ls}$. The properties of these estimators have been studied in **?** while for Euclidean data, there is a vast literature on leas square estimators for PLM see for example **?**, **?**, **?** and **?**.

3.2 Robust estimates

As in the Euclidean setting, the estimators introduced by **?** are a weighted average of the response variables. Hence, these estimates are very sensitive to large fluctuations of the variables so that, the final estimator of β can be seriously affected by anomalous data. To overcome this problem, **?** considered two families of robust estimators for the regression function when the explanatory variables t_i take values on a Riemannian manifold (M, γ). The first family combines the ideas of robust smoothing in Euclidean spaces with the kernel weights introduced in **?**. The second one generalizes to our setting the proposal given by **?**, who considered robust non parametric estimates using nearest neighbor weights, when the predictors t are on $I\!R^d$.

As in Bibi (1978), we consider a class of resistent estimates based on a three-step robust procedure under the partly linear model when some of the predictors take values on a Riemannian manifold. This approach does not require any moments to the errors, as is based on a natural extension of the conditional expectation to a setting where no moment conditions are required, studied in **?**. Denote by $\Psi : I\!R \rightarrow I\!R$ a strictly increasing, bounded and continuous function and by $F(y|t = \tau)$ and $F_j(x|t = \tau)$, the conditional distribution functions of y and x_j given $t = \tau$, respectively. Let $\phi_j(t)$ $0 \le j \le p$ be now any conditional location functionals related to a robust smoother. More precisely, for each $t \in M$ denote by $\phi_j(t)$ for $0 \le j \le p$ the solution of

$$E\left(\Psi\left(\frac{y - \phi_0(\tau)}{\sigma_0(\tau)}\right)\bigg| t = \tau\right) = 0 \quad \text{and} \quad E\left(\Psi\left(\frac{x_j - \phi_j(\tau)}{\sigma_j(\tau)}\right)\bigg| t = \tau\right) = 0 \quad \text{for } 1 \le j \le p \quad (4)$$

with $\sigma_0(\tau)$ and $\sigma_j(\tau)$ for $1 \le j \le n$ are robust measure of conditional scale respect to the conditional distribution of $y|t = \tau$ for $j = 0$ and $x_j|t = \tau$ for $1 \le j \le p$.

The three-step robust estimators are defined as follows:

Step 1: Estimate $\phi_j(t)$, $0 \le j \le p$ through a robust smoothing. Denote by $\widehat{\phi}_{j,\mathrm{R}}$ the obtained estimates and $\widehat{\phi}_{\mathrm{R}}(t) = (\widehat{\phi}_{1,\mathrm{R}}(t), \ldots, \widehat{\phi}_{p,\mathrm{R}}(t))^{\mathrm{T}}$.

Step 2: Estimate the regression parameter by applying a robust regression estimate to the residuals $y_i - \widehat{\phi}_{0,\mathrm{R}}(t_i)$ and $\mathbf{x}_i - \widehat{\phi}_{\mathrm{R}}(t_i)$. Denote by $\widehat{\beta}_{\mathrm{R}}$ the obtained estimator.

Step 3: Define the robust estimate of the regression function g as $\widehat{g}_{\mathrm{R}}(t) = \widehat{\phi}_{0,\mathrm{R}}(t) - \widehat{\beta}_{\mathrm{R}}^{\mathrm{T}}\widehat{\phi}_{\mathrm{R}}(t)$.

3

Note that in **Step 1**, the regression functions correspond to predictors taking values in a Riemannian manifold. Local $M-$type estimates $\widehat{\phi}_{0,\mathrm{R}}$ and $\widehat{\phi}_{j,\mathrm{R}}$ are defined in **?** as the solution of

$$\sum_{i=1}^{n} w_{n,h}(t,t_i)\Psi\left(\frac{y_i - \widehat{\phi}_{0,\mathrm{R}}(t)}{\sigma_{0,n}(t)}\right) = 0 \quad \text{and} \quad \sum_{i=1}^{n} w_{n,h}(t,t_i)\Psi\left(\frac{x_{ij} - \widehat{\phi}_{j,\mathrm{R}}(t)}{\sigma_{j,n}(t)}\right) = 0 \quad (5)$$

respectively, where the score function Ψ is strictly increasing, bounded and continuous and $\sigma_{0,n}(\tau)$ and $\sigma_{j,n}(\tau)$ $1 \le j \le p$ are local robust scale estimates. Possible choice for the score function Ψ can be the Huber or the bisquare Ψ-function. The local robust scale estimates $\sigma_{0,n}(\tau)$ and $\sigma_{j,n}(\tau)$ $1 \le j \le p$ can be taken as the local median of the absolute deviations from the local median (local MAD), *i.e.*, the MAD (see **?**) with respect to the distributions

$$F_n(y|t=\tau) = \sum_{i=1}^{n} w_{n,h}(\tau,t_i)I_{(-\infty,y]}(y_i) \quad \text{and} \quad F_{j,n}(x|t=\tau) = \sum_{i=1}^{n} w_{n,h}(\tau,t_i)I_{(-\infty,x]}(x_{ij}).$$

$$(6)$$

respectively.

In Step 2, the robust estimation of the regression parameter can be performed by applying to the residuals any of the robust methods proposed for linear regression. For example, we can consider M-estimates (**?**) or GM-estimators (**?**). On the other hand, high breakdown point estimates with high efficiency as MM-estimates may also be computed (**?** and **?**).

Throughout the paper, we consider $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ the solution of

$$\sum_{i=1}^{n} \psi_1\left((\widehat{r}_i - \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{R}})/s_n\right) w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right) \widehat{\boldsymbol{\eta}}_i = 0, \quad (7)$$

with s_n a robust consistent estimate of σ_ε, $\widehat{r}_i = y_i - \widehat{\phi}_{0,\mathrm{R}}(t_i)$, $\widehat{\boldsymbol{\eta}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t_i)$, ψ_1 a score function and w_1 a weight function. The zero of this equation can be computed iteratively using reweighting, as described for the location setting in [**?**, Chapter 2].

The estimator defined by **?** corresponds to the choice $\Psi(u) = u$ with the estimators of the conditional distribution based on kernel weights defined in (6). Therefore, if we considered the least square estimators of $\boldsymbol{\beta}$ in Step 2, we obtain the classical estimators proposed in **?**. On the other hand, when (M,γ) is $I\!\!R^d$ endowed with the canonical metric, the estimation procedure reduces to the proposal introduced in Bibi (1978). Details on the algorithm use to compute the robust non parametric estimators in Step 1 can be found in **?**.

4 Asymptotic results

The theorems of this Section study the asymptotic behavior of the regression parameter estimator of the PLM under standard conditions. Let U be an open set of M, we denote by $C^k(U)$ the set of k times continuously differentiable functions from U to $I\!\!R$. As in **?**, we assume that the image measure of P by t is absolutely continuous with respect to the Riemannian volume measure ν_γ and we denote by f its density on M with respect to ν_γ.

Let $\sigma_0(\tau)$ and $\sigma_j(\tau)$ for $1 \le j \le n$ be the MAD of the conditional distribution of $y_1|t=\tau$ for $j=0$ and $x_{1j}|t=\tau$ for $1 \le j \le n$, respectively.

4

4.1 Consistency

To derive strong consistency result of the estimate $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ of $\boldsymbol{\beta}$ defined in **Step 2** , we will consider the following set of assumptions.

$H1$. $\Psi : {\rm I\!R} \to {\rm I\!R}$ is an odd, strictly increasing, bounded and continuously differentiable function, such that $u\Psi'(u) \leq \Psi(u)$ for $u > 0$.

$H2$. $F(y|t = \tau)$ and $F_j(x|t = \tau)$ are symmetric around $\phi_0(\tau)$ and $\phi_j(\tau)$ and there are continuous functions of y and x for each τ.

$H3$. M_0 is a compact set on M such that:

   i) The density function f of t, is a bounded function such that $\inf_{\tau \in M_0} f(\tau) = A > 0$.

   ii) $\inf_{\substack{\tau \in M_0 \\ s \in M_0}} \theta_\tau(s) = B > 0$.

$H4$. The following equicontinuity condition holds

$$\forall \varepsilon > 0, \quad \exists \delta > 0 : |z - z'| < \delta \Rightarrow \sup_{s \in M_0} |G_s(z) - G_s(z')| < \varepsilon$$

when the function $G_s(z)$ equals $F(z|t = s)$ or $F_j(z|t = s)$ for $1 \leq j \leq p$.

$H5$. For any open set U_0 of M such that $M_0 \subset U_0$,

   i) f is of class C^2 on U_0.

   ii) $F(y|t = \tau)$ and $F_j(x|t = \tau)$ are uniformly Lipschitz in U_0, that is, there exists a constant $C > 0$ such that $|G_\tau(z) - G_s(z)| \leq C\, d_g(\tau, s)$ for all $\tau, s \in U_0$ and $z \in {\rm I\!R}$, when the function $G_s(z)$ equals $F(z|t = s)$ or $F_j(z|t = s)$ for $1 \leq j \leq p$.

$H6$. $K : {\rm I\!R} \to {\rm I\!R}$ is a bounded non negative Lipschitz function of order one, with compact support $[0, 1]$ satisfying $\int_{\boldsymbol{R}^d} \mathbf{u} K(\|\mathbf{u}\|) d\mathbf{u} = \mathbf{0}$ and $0 < \int_{\boldsymbol{R}^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|) d\mathbf{u} < \infty$.

$H7$. The sequence h_n is such that $h_n \to 0$ and $nh_n^d/\log n \to \infty$ as $n \to \infty$.

$H8$. The estimator $\sigma_{j,n}(\tau)$ of $\sigma_j(\tau)$ satisfy $\sigma_{j,n}(\tau) \xrightarrow{a.s.} \sigma_j(\tau)$ as $n \to \infty$ for all $\tau \in M_0$ and $0 \leq j \leq p$.

Remark 4.1.1. Assumption $H1$ is a standard condition in a robustness framework. Boundness of the score function allows to derive the weak continuity of the robust conditional functionals defined in (4) as shown in Theorem 2.2 of **?**. Differentiability of the score function is needed in order to obtain uniform consistency results over compact sets. Assumption $H2$ and the oddness of the score function allow to identify ϕ_j for $0 \leq j \leq p$. More precisely, these assumptions guarantee Fisher–Consistency, i.e. the definitions introduce in (4) coincide with the respective conditional expectations when they exist. The fact that $\theta_s(s) = 1$ for all $s \in M$ guarantees that $H3$ holds for a small compact neighborhood of s. $H4$ and $H5$ are needed in order to derive strong uniform consistency results (see **?**). Assumption $H6$ and $H7$, restricts the class

of kernel functions to be chosen and establishes conditions on the rate of convergence of the smoothing parameters, which are standard in non parametric regression. It is easy to see that Assumption $H8$ is satisfied, when we consider $\sigma_{j,n}(\tau)$ as the local median of the absolute deviations from the local median.

Remark 4.1.2. In order to obtain the consistency of the estimators, we will need that the regression parameter can be written as a functional on a distribution. More precisely, denote by P the distribution of $(r_i, \boldsymbol{\eta}_i)$ where $\boldsymbol{\eta}_i = \mathbf{x}_i - \boldsymbol{\phi}(t_i)$ and $r_i = y_i - \phi_0(t_i)$ and let $\boldsymbol{\beta}(H)$ be a regression functional, for the model $u = v^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$ where $(u, v) \sim H$ and u and ε are independent. Therefore, note that if $\widehat{r}_i = y_i - \widehat{\phi}_{0,\mathrm{R}}(t_i)$, $\widehat{\boldsymbol{\eta}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t_i)$ and $\widehat{P_n}(A) = \frac{1}{n}\sum_{i=1}^n I_A(\widehat{r}_i, \widehat{\boldsymbol{\eta}}_i)$, the robust estimator defined in (7) can be written as $\widehat{\boldsymbol{\beta}}_{\mathrm{R}} = \boldsymbol{\beta}(\widehat{P_n})$.

Theorem 4.1.1. *Let P and $\boldsymbol{\beta}(H)$ defined in Remark 4.1.2 and assume that $\boldsymbol{\beta}(H)$ is continuous in P and that it also provides Fisher-consistent estimates. Under assumptions $H1$ to $H8$, we have that*

a) $|\widehat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}| \xrightarrow{a.s.} 0$.

b) $\sup_{\tau \in M_0} |\widehat{g}_{\mathrm{R}}(\tau) - g(\tau)| \xrightarrow{a.s.} 0$.

4.2 Asymptotic distribution

In this Section, we assume that in **Step 2** of the estimation procedure, the robust estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ satisfies (7). More precisely, let ψ_1 be a score function and w_1 be a weight function, we will derive the asymptotic distribution of the regression parameter estimates $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ defined as a solution of

$$\sum_{i=1}^n \psi_1\left((\widehat{r}_i - \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{R}})/s_n\right) w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right) \widehat{\boldsymbol{\eta}}_i = 0,$$

with s_n a robust consistent estimate of σ_ε, $\widehat{r}_i = y_i - \widehat{\phi}_{0,\mathrm{R}}(t_i)$, $\widehat{\boldsymbol{\eta}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t_i)$. Denote by $\boldsymbol{\eta}_i = \mathbf{x}_i - \boldsymbol{\phi}(t_i)$ and $r_i = y_i - \phi_0(t_i)$. Note that $r_i - \boldsymbol{\eta}_i^{\mathrm{T}}\boldsymbol{\beta} = \varepsilon_i$.

To derive the asymptotic distribution of the regression parameter estimates, we will need the following set of assumptions.

A1. ψ_1 is an odd, bounded and twice continuously differentiable function with bounded derivatives ψ_1' and ψ_1'', such that the functions $u\psi_1'(u)$ and $u\psi_1''(u)$ are bounded.

A2. $E(w_1(\|\boldsymbol{\eta}_1\|)\,\boldsymbol{\eta}_1|t_1) = 0$, $E(w_1(\|\boldsymbol{\eta}_1\|)\,\|\boldsymbol{\eta}_1\|^2) < \infty$ and $A = E(\psi_1'(\varepsilon/\sigma_\varepsilon)w_1(\|\boldsymbol{\eta}_1\|)\,\boldsymbol{\eta}_1\boldsymbol{\eta}_1^{\mathrm{T}})$ is non singular.

A3. The function $w_1(u)$ is bounded, Lipschitz of order 1. Moreover, $\varphi(u) = w_1(u)u$ is also a bounded and continuously differentiable function with bounded derivative $\varphi'(u)$ such that $u\varphi'(u)$ is bounded.

A4. The functions $\phi_j(t)$ for $0 \le j \le p$ are continuous with ϕ_j' continuous in M.

6

A5. $\widehat{\phi}_j(t)$ the estimates of $\phi_j(t)$ for $0 \leq j \leq p$ have first continuous derivatives in M and

$$n^{1/4} \sup_{t \in M} |\widehat{\phi}_j(t) - \phi_j(t)| \xrightarrow{p} 0, \text{ for } 0 \leq j \leq p, \tag{8}$$

$$\sup_{t \in M} |\nabla \widehat{\phi}_j(t) - \nabla \phi_j(t)| \xrightarrow{p} 0, \text{ for } 0 \leq j \leq p. \tag{9}$$

where $\nabla \xi$ corresponds to the gradient of ξ.

A6. The estimator s_n of σ_ε satisfies $s_n \xrightarrow{p} \sigma_\varepsilon$ as $n \to \infty$.

Theorem 4.2.1. Under the assumptions $A1$ to $A6$ we have that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, \sigma_\varepsilon^2 A^{-1} \Sigma A^{-1}),$$

where A is defined in $A2$ and $\Sigma = E(\psi_1^2(\varepsilon/\sigma_\varepsilon))E(w_1^2(\|\boldsymbol{\eta}_1\|)\,\boldsymbol{\eta}_1\boldsymbol{\eta}_1^{\mathrm{T}})$.

Remark 4.2.1. To prove the previous result, we will need an inequality for the covering numbers of the class $\mathcal{F}(M) = \{\xi \in \mathcal{C}^1(M) : \|\xi\|_\infty \leq 1 \quad \|\nabla \xi\|_\infty \leq 1\}$. In the Appendix, we include some results related to the covering number on a Riemannian manifold.

5 Simulation study

In this section, we report the results of a simulation study designed to evaluate the performance of the robust procedure introduced in Section 3. The main objective of this study is to compare the behavior of the classical and robust estimators for normal and contaminated samples. We consider the cylinder endowed with the metric induced by the canonical metric of $I\!\!R^3$. Due to of the computational complexity of the robust procedure, we performed 500 replications of independent samples of size $n = 200$. In the smoothing procedure, the kernel was taken as the quadratic kernel $K(t) = (15/16)(1-t^2)^2 I(|x| < 1)$. The local M−estimate was computed with bisquare score function, with constant 4.685, which gives a 95% efficiency. As initial estimate in the iterative procedure to compute the local M−estimate, we have considered the local median. For the regression parameter, we have considered a GM-estimators (7) with score function on the residuals $\psi_1(r) = \psi_{\mathrm{H},c}(r) = \max(-c, \min(r, c))$, i.e, the Huber function with tunning constant c. In the simulation study, we considered two different tunning constants $c = 1.6$ and $c = 1.7$ and weight function w_1

$$w_1(\eta) = W\left[((\eta - \mu_\eta)/\sigma_\eta)^2\right] \tag{10}$$

where $W(t) = \psi_{\mathrm{H},\chi_{1,0.975}}(t)/t$ while $\mu_\eta = \underset{1 \leq i \leq n}{\mathrm{median}}(\widehat{\boldsymbol{\eta}}_i)$ and $\sigma_\eta = \underset{1 \leq i \leq n}{\mathrm{mad}}(\widehat{\boldsymbol{\eta}}_i)/0.6754$ with $\widehat{\boldsymbol{\eta}}_i = \mathbf{x}_i - \widehat{\phi}_{\mathrm{R}}(t_i)$.

In the next section, we describe the robust cross validation procedure used in order to obtain the robust estimates. To compute the classical estimators, the bandwidth was selected using the classical cross validation described in **?**. The distance d_γ and the volume density function for the cylinder were computed in **?** and **?**. We consider the following model:

The variables (y_i, x_i, t_i) for $1 \leq i \leq n$ are generated as

$$y_i = 2\, x_i + (t_{1i} + t_{2i} - t_{3i})^2 + \varepsilon_i \quad \text{and} \quad x_i = \sin(2t_{3i}) + \eta_i$$

where $t_i = (t_{1i}, t_{2i}, t_{3i}) = (\cos(\theta_i), \sin(\theta_i), s_i)$ with the variables θ_i following a uniform distribution in $(0, 2\pi)$ and the variables s_i uniform in $(0, 1)$, i.e., t_i have support in the cylinder with radius 1 and height between $(0, 1)$.

The non contaminated case, denoted C_0 corresponds to independent errors ε_i and η_i normally distributed with mean 0 and standard deviation 1 and 0.05, respectively. Besides, the so-called contaminations C_1 and C_2, correspond to selecting a distribution in a neighborhood of the central normal distribution and are defined as $\varepsilon \sim 0.9N(0, 1) + 0.1N(0, 25)$ and $\varepsilon \sim 0.9N(0, 1) + 0.1N(5, 0.25)$, respectively. Also we consider the contamination denoted C_3, where we introduce artificially 10 observations of the variables x equal to 5 but we did not change the response variables and the covariates t.

The contamination C_1 which corresponds to inflating the errors will affect the variance of the regression estimates, while C_2 is an asymmetric contamination. The contamination C_3 allows to study the behavior of the estimators under the presence of high leverage points.

5.1 Bandwidth Selection

To select the smoothing parameter there exist two commonly used approaches: L^2 cross-validation and plug-in methods. However, these procedures may not be robust. Their sensitivity to anomalous data was discussed by several authors, see for example **?**, **?**, **?**, **?** and **?**. Under a non parametric regression model with carriers in an Euclidean space for spline-based estimators, **?** introduced a robust cross-validation criterion to select the bandwidth parameter. Robust cross-validation selectors for kernel M-smoothers were considered in **?**, **?** and **?**, under a fully non parametric regression model. In the Euclidean setting, for partly linear model, a robust plug-in procedure was studied in **?** while for dependent observations, a robust cross-validation criterion was considered in **?**. When the variables belong in a Riemannian manifold, a robust cross validation procedure was discussed in **?** under a fully non parametric regression model, while a classical cross-validation procedure under a partly linear models was considered in **?**.

We describe a robust cross-validation method to select the bandwidth in the case of partly linear models that robustifies the proposal given in **?** and generalizes the procedure given in **?**. The robust cross-validation method constructs an asymptotically optimal data-driven bandwidth, and thus adaptive data-driven estimators, by minimizing

$$RCV(h) = \sum_{i=1}^{n} \mu_n^2(\widehat{\varepsilon}_i(h)) + \sigma_n^2(\widehat{\varepsilon}_i(h)),$$

with $\widehat{\varepsilon}_i(h) = y_i - \widehat{\phi}_{0,-i,h}(t_i)) - (\mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{-i,h}(t_i))^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}$; $\widehat{\phi}_{0,-i,h}(t)$ and $\widehat{\boldsymbol{\phi}}_{-i,h}(t) = (\widehat{\phi}_{1,-i,h}(t), \ldots, \widehat{\phi}_{p,-i,h}(t))$ denote the robust non parametric estimators computed with bandwidth h using all the data expect the i−th observation and $\widetilde{\boldsymbol{\beta}}$ estimate the regression parameter by applying a robust

regression estimate to the residuals $y_i - \widehat{\phi}_{0,-i,h}(t_i)$ and $\mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{-i,h}(t_i)$. Besides, μ_n and σ_n^2 denote robust estimators of location and scale, respectively.

In the simulation study and in the real data example, we have consider μ_n as the median and σ_n as the Huber $\tau-$scale estimator. Also, the search for the bandwidth parameter was performed over the following values of bandwidths $0.1, 0.25, 1, 1.5, 2.5, 4, 5, 6.5$.

The asymptotic properties of data-driven estimators require further careful investigation and are beyond the scope of this paper.

5.2 Simulation results

Table 5.1 shows 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 500 replications for the considered model. We denote with ls, R, 1.6 and R, 1.7 the classical and robust estimators with tuning constants $c = 1.6$ and $c = 1.7$, respectively. Figure 5.1 shows the boxplot of the regression parameter estimators. Since the results are very similar for the robust estimators using different tuning constants, in the figure, we only report the results for the robust estimators for $c = 1.6$.

	mean($\widehat{\boldsymbol{\beta}}_{ls}$)	sd($\widehat{\boldsymbol{\beta}}_{ls}$)	MSE($\widehat{\boldsymbol{\beta}}_{ls}$)	MSE($\widehat{g}_{ls}$)
C_0	2.0732	0.1445	0.0262	0.2396
C_1	1.8789	1.7592	3.1095	20.4485
C_2	1.8722	1.7975	3.2475	45.9719
C_3	0.2806	0.0804	2.9627	1.4988
	mean($\widehat{\boldsymbol{\beta}}_{R,1.6}$)	sd($\widehat{\boldsymbol{\beta}}_{R,1.6}$)	MSE($\widehat{\boldsymbol{\beta}}_{R,1.6}$)	MSE($\widehat{g}_{R,1.6}$)
C_0	2.0646	0.1524	0.0274	0.2431
C_1	2.0198	0.2303	0.0534	0.4897
C_2	2.0109	0.2540	0.0646	1.3580
C_3	1.9563	0.2318	0.0557	0.3635
	mean($\widehat{\boldsymbol{\beta}}_{R,1.7}$)	sd($\widehat{\boldsymbol{\beta}}_{R,1.7}$)	MSE($\widehat{\boldsymbol{\beta}}_{R,1.7}$)	MSE($\widehat{g}_{R,1.7}$)
C_0	2.0458	0.1569	0.0267	0.2717
C_1	2.0151	0.2375	0.0566	0.3794
C_2	1.9997	0.3082	0.0950	2.2173
C_3	1.9468	0.2272	0.0544	0.3652

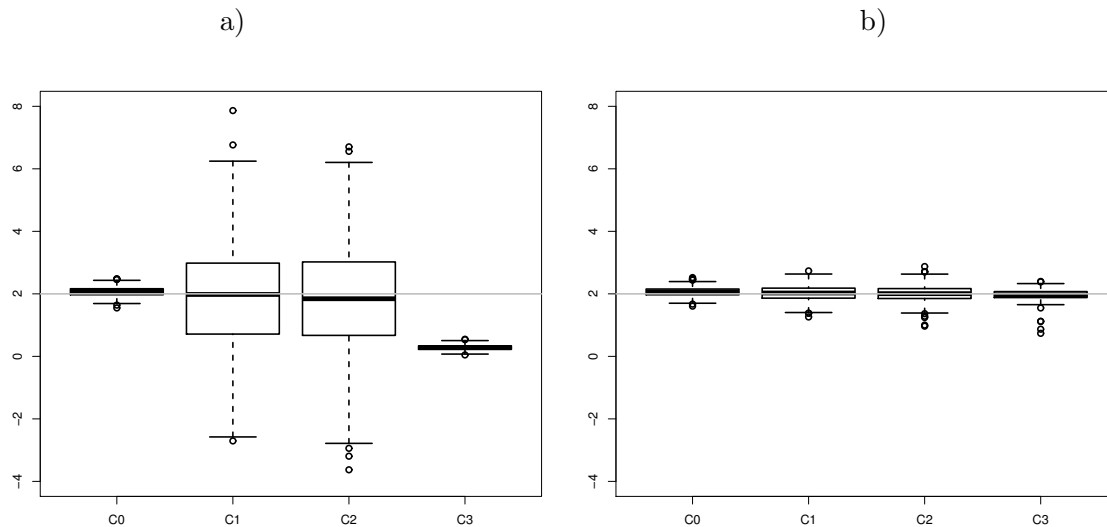Table 5 Performance of regression parameter and the regression functions under the different contaminations.

Figure 5. Boxplot of a) $\widehat{\boldsymbol{\beta}}_{ls}$ the classical estimators and b) $\widehat{\boldsymbol{\beta}}_{R,1.6}$ the robust estimators with tuning constant $c = 1.6$ under the different contaminations.

The simulation results confirm the inadequate behavior of the classical estimators under the considered contaminations. The robust estimators of the regression function introduced in this paper show only a small lack of efficiency under normality. Under C_1, C_2 and C_3, the results obtained with the classical estimators are not reliable giving larger mean square errors obtained that those the robust procedure. In Table 4.1, we can observe that the contamination C_2 not only affect the classical estimators of the regression function, but also for the robust ones for both tuning constants. In particular under this contamination, the mean square error both for the classical regression estimator or the classical regression function estimator are of order 125 and 190 times larger than in the non contaminated setting. However, for the robust estimators with $c = 1.6$, the mean square error for the regression estimator and the regression function estimator increased 2.5 and 5.5 times than the non contaminated case, respectively. Under the contamination C_3, we can observe how the high leverage points affect the bias of the classical regression estimators. These extreme behaviors of the classical estimators show their inadequacy when one suspects that the sample can contains outliers. The results for the robust estimators using different tuning constants are quite similar.

6 Real Example

The solar insolation is the amount of electromagnetic energy or solar radiation incident on the surface of the earth. This variable measures the duration of sunlight in seconds. In the automatic stations, the World Meteorological Organization defines insolation as the sum of time intervals in which the irradiance exceeds the threshold of 120 watts per square meter. The irradiance is direct radiation normal or perpendicular to the sun on Earth's surface. The values of the insolation in a particular location depend of the weather conditions and the sun's

position on the horizon. For example, the presence of clouds increases the absorption, reflection and dispersion of the solar radiation. Desert areas, given the lack of clouds, have the highest values of insolation on the planet. More details about insolation can be seen in **?**.

As commented above, the isolation is related with the weather conditions. In particular, to illustrate the proposed estimators, we will analyze the relation between the insolation, the humidity, the direction and the speed of the wind. We consider a data set available at http://meteo.navarra.es/. This data consists on the daily average of relative humidity, speed and direction of the wind and the insolation. The direction's wind was measured with the point zero in the north direction and the wind's speed was measured in meters per second. The recorded data were measured daily in the automatic meteorologic station of Pamplona-Larrabide GN, in Navarra, Spain during the year 2004. In our study, we selected a random sample from this dataset.

In Figure 5.1, we can see that the humidity and the insolation follow a lineal relation except for the points contained in the ellipse on the left of the plot. Therefore, we consider a partly lineal model to explain the insolation, as a linear function of the humidity and a non parametric function of the speed and direction of the wind. Note that, the variables corresponding to the wind to be modeled non parametrically, belong to a cylinder. In the smoothing procedure, we choose the quadratic kernel $K(t) = (15/16)(1 - t^2)^2 I(|x| < 1)$. The robust estimators of the parameter and the regression function were computed with the same scores and the weight functions that we consider in the simulation study. For the robust estimators, the bandwidth was selected using the robust cross validation procedure described in Section 4.1. For the classical estimators, a least square cross validation described in **?** was considered.
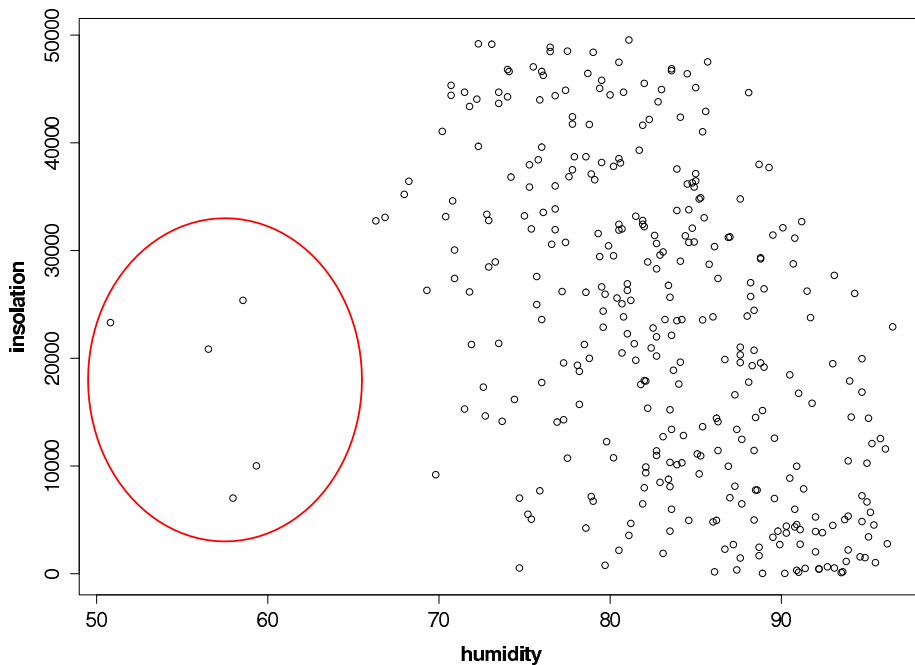

Figure 6 Scatter plot between the insolation and humidity. The dots in the ellipse correspond to the potential outliers.

11

In a first step, we apply the classical and robust methods to obtain an estimator of the regression parameter using all the data. The obtained results are $\widehat{\boldsymbol{\beta}}_{ls} = -1032.869$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{R}} = -1246.856$. Also, based in the asymptotic results obtained in Theorem 3.2.1, we calculate confidence intervals with level 0.95 estimating the unknown quantities. The result of the classical confidence interval is $CCI_{0.05}(\beta) = (-1229.9451 - 835.7935)$ while that based in the robust estimation $RCI_{0.05}(\beta) = (-1453.658, -1040.053)$. We observe a shift between the classical and robust intervals that may be due to the effect of the observations with low values of humidity. For that reason, we compute the classical estimator using all the data except the potential outliers obtaining $\widehat{\boldsymbol{\beta}}_{ls} = -1294.620$ with related confidence interval $CCI_{0.05}(\beta) = (-1502.983, -1086.257)$ giving values closer to those of the robust procedure. It is clear that, if we estimate the regression parameter with the classical approach when the dataset have outliers, the conclusions can be misleading. For example, with the classical estimator computed with all the data, the hypothesis that $\beta = -1000$ is not rejected, while the conclusions with the classical estimator without the outliers or the robust estimators with all the data are reversed, rejecting the null hypothesis.

Acknowledgments

A   Appendix

A.1   Proof of Theorem 4.1.1.

a) Since Remark 4.1.2, it is suffice to prove that $\Pi(\widehat{P_n}, P) \xrightarrow{a.s.} 0$ where Π stands for the Prohorov distance. Thus, we will show that for any bounded and continuous function $f : I\!\!R^{p+1} \to I\!\!R$ we have that $|E_{\widehat{P_n}} f - E_P f| \xrightarrow{a.s.} 0$.

Given $\varepsilon > 0$, we have the bound

$$
\begin{aligned}
|E_{\widehat{P_n}} f - E_P f| \;\leq\; & \frac{1}{n}\sum_{i=1}^{n} |f(r_i + (\phi_0(t_i) - \widehat{\phi}_0(t_i)), \boldsymbol{\eta}_i + (\boldsymbol{\phi}(t_i) - \widehat{\boldsymbol{\phi}}(t_i))) - f(r_i, \boldsymbol{\eta}_i)| I_C(r_i, \boldsymbol{\eta}_i, t_i) \\
+ \; & \frac{1}{n}\sum_{i=1}^{n} I_{C^c}(r_i, \boldsymbol{\eta}_i, t_i) \|f\|_\infty
\end{aligned}
$$

where $C_1 \subset I\!\!R^{p+1}$ and $M_0 \subset M$ are compact sets that $P(C) > 1 - \varepsilon/(4\|f\|_\infty)$ with $C = C_1 \times M_0$.

Using Theorem 3.3 of **?**, we have that

$$
\sup_{t \in M_0} |\widehat{\phi}_{j,\mathrm{R}}(t) - \phi_j(t)| \xrightarrow{a.s.} 0 \tag{A.1}
$$

for $0 \leq j \leq p$. From this fact and the Strong Law of Large Numbers, we have that there exists a set $\aleph \subset \Omega$ such that $P(\aleph) = 0$ and for any $\omega \notin \aleph$ we have that (A.1) holds and

$$\frac{1}{n} \sum_{i=1}^{n} I_{C^c}(r_i, \boldsymbol{\eta}_i, t_i) \to P(C^c).$$

Let $\bar{C}_1$ the closure of a neighborhood of radius 1 of C_1. The uniform continuity of f on $\bar{C}_1$ implies that there exists δ such that $\max_{1 \leq j \leq p+1} |u_j - u_i|$, $u, v \in \bar{C}_1$ entails $|f(u) - f(v)| \leq \frac{\varepsilon}{2}$. Thus, we have that for $\omega \notin \aleph$ and n large enough $\max_{0 \leq j \leq p} \sup_{t \in M_0} |\widehat{\phi}_{j,\mathrm{R}}(t) - \phi_j(t)| < \delta$ so that, for $1 \leq i \leq n$, we obtain that

$$|f(r_i + (\phi_0(t_i) - \widehat{\phi}_0(t_i)), \boldsymbol{\eta}_i + (\boldsymbol{\phi}(t_i) - \widehat{\boldsymbol{\phi}}(t_i))) - f(r_i, \boldsymbol{\eta}_i)| \leq \frac{\varepsilon}{2}.$$

concluding the proof.

b) Follows inmediately from (A.1) and a). $\square$

A.2 Entropy number

The main objective of this Section is to obtain an upper-bound to the entropy number of the class of functions $\mathcal{F}(M) = \{\xi \in \mathcal{C}^1(M) : \|\xi\|_\infty \leq 1 \quad \|\nabla \xi\|_\infty \leq 1\}$. The covering number $N(\delta, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls, $\{\xi : \|\xi - \eta\| < \delta\}$ of radius δ needed to cover the set $\mathcal{F}$. The entropy number is the logarithm of the covering number. This upper-bound will be used to obtain the asymptotic distribution of the regression parameter. Several authors have studied bounds to the covering numbers for different sets, see for example **?**, **?** and **?**. In particular, **?** obtained an upper-bound for the covering number of $\mathcal{F}(M)$ when M is a bounded, convex subset of $I\!\!R^d$. For the convenience of the reader, we have included the following remark (see **?**).

Remark A.1. Let $N(\delta)$ be the minimal number of balls with radius δ needed to cover (M, γ). A δ-filling in M is a maximal family of pairwise disjoint open balls of radius δ contained in M. We denote by $D(\delta)$ the maximum number of such balls. Is easy to see that $N(2\delta) \leq D(\delta)$. Let $diam_{(M,\gamma)}$ be the diameter of (M, γ) and consider $\kappa \in I\!\!R$ such that $Ricc_{(M,\gamma)} \geq (d-1)\kappa$ where $Ricc_{(M\gamma)}$ is the Ricci curvature and d the dimension of M. For example, if γ is an Einstein metric's with scalar curvature $2(d-1)\kappa$ then the inequality is attained. Note that if $\kappa > 0$ since Myers's Theorem **?**, (M, γ) is a compact manifolds with $diam_{(M,\gamma)} \leq \pi/\sqrt{\kappa}$. Since M is compact there exists κ with this property. Denote by $V^\kappa(r)$ the volume of a ball of radius r in a complete, simply connected Riemannian manifold with constant curvature κ. By the Theorem of Bishop (see **?**) we know that $\frac{Vol(B(x,r))}{V^\kappa(r)}$ is a non increasing function where $B(x,r) = \{z \in M : d_\gamma(x,z) \leq r\}$ is the geodesic ball centered in x with radius r. Note that, M is the closure of $B(x, diam_{(M,\gamma)})$ for any $x \in M$. If $\{B(a_1, \frac{\delta}{2}), \dots, B(a_D, \frac{\delta}{2})\}$ with $D = D(\frac{\delta}{2})$ is a $\frac{\delta}{2}$-filling then,

$$N(\frac{\delta}{2}) \leq \frac{Vol(M)}{\inf_{1 \leq i \leq D} Vol(B(a_i, \frac{\delta}{2}))} \leq \frac{V^\kappa(diam_{(M,\gamma)})}{V^\kappa(\frac{\delta}{2})}.$$

Therefore $N(\delta) \leq C(diam_{(M,\gamma)}, \kappa)\delta^{-d}$.

Lemma A.1. Let $\mathcal{F}(M) = \{\xi \in \mathcal{C}^1(M) : \|\xi\|_\infty \leq 1 \quad \|\nabla\xi\|_\infty \leq 1\}$, then the covering number for the supremum norm of $\mathcal{F}(M)$ that we denote by $N(\delta, \mathcal{F}(M), \|\cdot\|_\infty)$ satisfies that $\log N(\delta, \mathcal{F}(M), \|\cdot\|_\infty) < A\delta^{-d}$.

Proof of Lemma A.1. Let $\mathcal{A} = \{B(a_1, \delta), \ldots, B(a_N, \delta)\}$ be a covering of M by open balls of radius δ. By the remark above, we may assume that $N \leq C(diam_{(M,\gamma)}, \kappa)\delta^{-d}$. Also, we can choose the covering $\mathcal{A}$ such that $B(a_i, \delta) \cap B(a_{i+1}, \delta) \neq \emptyset$ for $1 \leq i \leq N - 1$ and $a_i \neq a_j$ for $1 \leq i, j \leq N$. Let $\xi \in \mathcal{F}(M)$, we define the function $\widetilde{\xi} = \sum_{i=1}^N \delta \left[\frac{\xi(a_i)}{\delta}\right] I_{D_i}$ where $D_1 = B(a_1, \delta)$, $D_i = B(a_i, \delta) \setminus \cup_{j=1}^{i-1} B(a_j, \delta)$ and $[a]$ denotes the integer part of a.

Let $x \in M$ and $1 \leq k \leq N$ such that $x \in D_k$, then we have that $|\widetilde{\xi}(x) - \xi(x)| \leq |\widetilde{\xi}(x) - \xi(a_k)| + |\xi(a_k) - \xi(x)|$. Since $\widetilde{\xi}(a_k) = \widetilde{\xi}(x)$ and $\xi(a_k) = \widetilde{\xi}(a_k) + \delta(\frac{\xi(a_k)}{\delta} - [\frac{\xi(a_k)}{\delta}]) = \widetilde{\xi}(a_k) + \delta B$ with $0 \leq B < 1$ and the fact that $\|\nabla\xi\| \leq 1$, we have that $|\widetilde{\xi}(x) - \xi(x)| \leq 2\delta$.

For the first value $\widetilde{\xi}(a_1)$ of a generic function ξ, we have $4[\frac{1}{\delta}] + 1$ possibilities since $|\widetilde{\xi}(a_1)| \leq 1$. Using that,

$$|\widetilde{\xi}(a_k) - \widetilde{\xi}(a_{k-1})| \leq |\widetilde{\xi}(a_k) - \xi(a_k)| + |\xi(a_k) - \xi(a_{k-1})| + |\xi(a_{k-1}) - \widetilde{\xi}(a_{k-1})| \leq 6\delta.$$

We get that, for each value of $\widetilde{\xi}(a_{k-1})$ we only have 13 possibilities to choose $\widetilde{\xi}(a_k)$. Then, it is easy to verify that

$$N(2\delta, \mathcal{F}(M), \|\cdot\|_\infty) \leq (4[\frac{1}{\delta}] + 1)13^N.$$

concluding the proof. $\square$

Remark A.2. Since $N(\delta, \mathcal{F}(M), L^2(Q)) \leq N(\delta, \mathcal{F}(M), \|\cdot\|_\infty)$, Lemma A.1 entails that the covering number of $\mathcal{F}(M)$ satisfies, $\log N(\delta, \mathcal{F}(M), L^2(Q)) < A\delta^{-d}$.

A.3   Proof of Theorem 4.2.1.

Using a Taylor expansion of order one around $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$, we get that $S_n = A_n(\widehat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta})$ with

$$S_n = \frac{1}{n}\sum_{i=1}^n \psi_1\left((\widehat{r}_i - \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\boldsymbol{\beta})/s_n\right) w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)\widehat{\boldsymbol{\eta}}_i$$

$$A_n = \frac{1}{n}\sum_{i=1}^n \psi_1'\left((\widehat{r}_i - \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}})/s_n\right) w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)\widehat{\boldsymbol{\eta}}_i\widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}.$$

where $\widetilde{\boldsymbol{\beta}}$ is an intermediate point between $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$. Analogous arguments to those used in Lemma 2 in Bibi (1978) allow to show that $A_n \xrightarrow{p} A$ where A is defined in $A2$.

Since $\frac{\sqrt{n}}{n}\sum_{i=1}^n \psi_1(\varepsilon_i/\sigma_\varepsilon) w_1(\|\boldsymbol{\eta}_i\|)\boldsymbol{\eta}_i$ is asymptotically normally distributed with covariance $\boldsymbol{\Sigma}$, it will enough to show that

$$\sqrt{n}\quad [S_n - \frac{1}{n}\sum_{i=1}^n \psi_1(\varepsilon_i/s_n) w_1(\|\boldsymbol{\eta}_i\|)\boldsymbol{\eta}_i] \xrightarrow{p} 0, \tag{A.2}$$

$$\sqrt{n} \quad [\frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\varepsilon_i/s_n\right)w_1\left(\|\boldsymbol{\eta}_i\|\right)\boldsymbol{\eta}_i - \frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\varepsilon_i/\sigma_\varepsilon\right)w_1\left(\|\boldsymbol{\eta}_i\|\right)\boldsymbol{\eta}_i] \xrightarrow{p} 0. \qquad \text{(A.3)}$$

We first prove (A.2). Using a Taylor expansion of order two, we have that the following decomposition.

$$\sqrt{n}[S_n - \frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\varepsilon_i/s_n\right)w_1\left(\|\boldsymbol{\eta}_i\|\right)\boldsymbol{\eta}_i] \quad = \quad \sum_{i=1}^{5}S_{ni}$$

with

$$S_{n1} \quad = \quad \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\psi_1'\left(\varepsilon_i/s_n\right)[\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}(t_i)\boldsymbol{\beta} - \widehat{\gamma}_0(t_i)]w_1\left(\|\boldsymbol{\eta}_i\|\right)\boldsymbol{\eta}_i$$

$$S_{n2} \quad = \quad \frac{s_n\sqrt{n}}{n}\sum_{i=1}^{n}\psi_1\left(\varepsilon_i/s_n\right)[w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)\widehat{\boldsymbol{\eta}}_i - w_1\left(\|\boldsymbol{\eta}_i\|\right)\boldsymbol{\eta}_i]$$

$$S_{n3} \quad = \quad \frac{s_n\sqrt{n}}{n}\sum_{i=1}^{n}[\psi_1\left(\widehat{r}_i - \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\boldsymbol{\beta}/s_n\right) - \psi_1\left(\varepsilon_i/s_n\right)]w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)[\widehat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i]$$

$$S_{n4} \quad = \quad \frac{\sqrt{n}}{2n}\sum_{i=1}^{n}\psi_1''\left(\varsigma_i/s_n\right)[\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}(t_i)\boldsymbol{\beta} - \widehat{\gamma}_0(t_i)]^2 w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)\boldsymbol{\eta}_i$$

$$S_{n5} \quad = \quad \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\psi_1\left(\varepsilon_i/s_n\right)[\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}(t_i)\boldsymbol{\beta} - \widehat{\gamma}_0(t_i)][w_1\left(\|\widehat{\boldsymbol{\eta}}_i\|\right) - w_1\left(\|\boldsymbol{\eta}_i\|\right)]\boldsymbol{\eta}_i,$$

where $\widehat{\gamma}_j(t) = \widehat{\phi}_j(t) - \phi_j(t)$ for $0 \leq j \leq n$ and $\widehat{\boldsymbol{\gamma}}(t) = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_n)^{\mathrm{T}}$. By $A3$, $A5$ and $A6$ is easy to see that $\|S_{in}\| \xrightarrow{p} 0$ for $i = 3, 4, 5$.

Let

$$\mathcal{J}_{1n}^{(j)}(\sigma, \xi) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n}f_{1,\sigma,\xi}^{(j)}(r_i, \boldsymbol{\eta}_i, t_i)$$

$$= \quad \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{r_i - \boldsymbol{\eta}_i^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}\right)\xi(t_i)w_1\left(\|\boldsymbol{\eta}_i\|\right)(\boldsymbol{\eta}_i)_j$$

$$\mathcal{J}_{2n}^{(j)}(\sigma, \boldsymbol{\xi}) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n}f_{2,\sigma,\boldsymbol{\xi}}^{(j)}(r_i, \boldsymbol{\eta}_i, t_i)$$

$$= \quad \frac{\sigma\sqrt{n}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{r_i - \boldsymbol{\eta}_i^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}\right)[w_1\left(\|\boldsymbol{\eta}_i + \boldsymbol{\xi}\|\right)(\boldsymbol{\eta}_i + \boldsymbol{\xi}(t_i))_j - w_1\left(\|\boldsymbol{\eta}_i\|\right)(\boldsymbol{\eta}_i)_j]$$

Therefore, it remains to show that $\mathcal{J}_{1n}^{(j)}(s_n, \widehat{\gamma}_s) \xrightarrow{p} 0$ and $\mathcal{J}_{1n}^{(j)}(s_n, \widehat{\boldsymbol{\gamma}}) \xrightarrow{p} 0$ for $0 \leq j, s \leq p$. From now on, we will omitted the superscript j for the sake of simplicity.

Let $\mathcal{F}(M) = \{\xi \in \mathcal{C}^1(M) : \|\xi\|_\infty \leq 1 \ \|\nabla\xi\|_\infty \leq 1\}$ and consider the classes of functions

$$\mathcal{F}_1 \quad = \quad \{f_{1,\sigma,\xi}(r, \boldsymbol{\eta}, t) \quad \sigma \in (\sigma_\varepsilon/2, 2\sigma_\varepsilon) \quad \xi \in \mathcal{F}(M)\}$$
$$\mathcal{F}_2 \quad = \quad \{f_{2,\sigma,\boldsymbol{\xi}}(r, \boldsymbol{\eta}, t) \quad \sigma \in (\sigma_\varepsilon/2, 2\sigma_\varepsilon) \quad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_p), \ \xi_s \in \mathcal{F}(M)\}$$

Note that, the independence of ε_i and $(\mathbf{x}_i, t_i)$, $A2$ and the fact that the errors ε have symmetric distribution imply that $E(f(r_i, \boldsymbol{\eta}_i, t_i)) = 0$ for any $f \in \mathcal{F}_1 \cup \mathcal{F}_2$. As in Bibi (1978), it is easy to see that the covering number of the classes $\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy

$$N(C_1\epsilon, \mathcal{F}_1, L^2(Q)) \leq N(\epsilon, \mathcal{F}(M), L^2(Q)) \, N(\varepsilon, (\sigma_\varepsilon/2, 2\sigma_\varepsilon), |\cdot|)$$

$$N(C_2\epsilon, \mathcal{F}_2, L^2(Q)) \leq N^p(\epsilon, \mathcal{F}(M), L^2(Q)) \, N(\varepsilon, (\sigma_\varepsilon/2, 2\sigma_\varepsilon), |\cdot|)$$

where Q is any probability measure. From Remark A.2 we get that the covering number of $\mathcal{F}(M)$ satisfies that $\log N(\epsilon, \mathcal{F}(M), L^2(Q)) < A\varepsilon^{-d}$. Therefore, the classes $\mathcal{F}_1$ and $\mathcal{F}_2$ have finite uniform-entropy. For $0 < \delta < 1$, consider the subclasses $\mathcal{F}_{1,\delta}$ and $\mathcal{F}_{2,\delta}$ of $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively, defined by,

$$\begin{aligned}
\mathcal{F}_{1,\delta} &= \{f \in \mathcal{F}_1 \quad \xi \in \mathcal{F}(M), \; \|\xi\|_\infty < \delta\} \\
\mathcal{F}_{2,\delta} &= \{f \in \mathcal{F}_2 \quad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_p), \; \xi_s \in \mathcal{F}(M), \; \|\xi_s\|_\infty < \delta\}
\end{aligned}$$

For any $\epsilon > 0$, let $0 < \delta < 1$ from $A5$ and $A6$, we obtain that for n large enough $P(s_n \in (\sigma_\varepsilon/2, 2\sigma_\varepsilon)) > 1 - \delta/2$ and $P(\widehat{\gamma}_s \in \mathcal{F}(M)$ and $\|\widehat{\gamma}_s\|_\infty < \delta) > 1 - \delta/2$ for $0 \leq s \leq p$.

Then, the maximal inequality for covering numbers entails that for $0 \leq s \leq p$

$$\begin{aligned}
P(|\mathcal{J}_{1n}(s_n, \widehat{\gamma}_s)| > \epsilon) &\leq P(|\mathcal{J}_{1n}(s_n, \widehat{\gamma}_s)| > \epsilon; \; s_n \in (\sigma_\varepsilon/2, 2\sigma_\varepsilon); \; \widehat{\gamma}_s \in \mathcal{F}(M) \text{ and } \|\widehat{\gamma}_s\|_\infty < \delta) + \delta \\
&\leq P\left(\sup_{f \in \mathcal{F}_{1,\delta}} \left| \frac{\sqrt{n}}{n} \sum_{i=1}^n f(r_i, \boldsymbol{\eta}_i, t_i) \right| > \epsilon \right) + \delta \\
&\leq \frac{1}{\epsilon} E\left(\sup_{f \in \mathcal{F}_{1,\delta}} \left| \frac{\sqrt{n}}{n} \sum_{i=1}^n f(r_i, \boldsymbol{\eta}_i, t_i) \right| \right) + \delta \\
&\leq \frac{1}{\epsilon} \mathcal{G}(\delta, \mathcal{F}_1) + \delta
\end{aligned}$$

where $\mathcal{G}(\delta, \mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L^2(Q))} d\epsilon$. Using that $\mathcal{F}_1$ satisfies the uniform–entropy condition we get that $\lim_{\delta \to 0} \mathcal{G}(\delta, \mathcal{F}_1) = 0$, therefore $S_{1n} \xrightarrow{p} 0$. Similarly arguments considered for $\mathcal{J}_{2n}(s_n, \widehat{\gamma})$ and the class $\mathcal{F}_2$ to obtain that $S_{2n} \xrightarrow{p} 0$.

The proof of (A.3), follows using analogous arguments that those considered in (A.2).$\square$

References

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

Bianco, A. and Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Plann. Inference.* **122,** 229–252.