

Robust estimators in partly linear regression models on Riemannian manifolds

Guillermo Henry and Daniela Rodriguez

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina

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 predictor variables take values on a Riemannian manifold. We obtain the consistency and the asymptotic normality of the proposed estimators. Simulations and application to real data show the performance of our proposal under small samples and contamination.

Key words and phrases: Nonparametric 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 nonparametrically on some others. To be more precise, assume that we have a response $y_i \in \mathbb{R}$ and covariates (\mathbf{x}_i, t_i) such that $\mathbf{x}_i \in \mathbb{R}^p, t_i \in [0, 1]$ satisfying

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

where the errors ε_i are independent and independent of (\mathbf{x}_i^T, t_i) . Since the introductory work by [10], the partly linear models have become an important tool in the modeling of econometric or biometric data, combining the flexibility of nonparametric models and the simple interpretations of the linear ones. 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.

In a recent work [11], PLM are studied when the explanatory variables takes values on a Riemannian manifold and the potential application of this model in an environment problem are explored. Unfortunately, as we will see in Section 4, 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 derive robust estimators to fit PLM models 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 manifolds.

This paper is organized as follows. In Section 2, we give a brief summary of the classical proposal of estimation for this model and we introduce the robust estimates. In Section 3, we study the consistency and the asymptotic distribution of the regression parameter under regular assumptions on the bandwidth sequence. Section 4 includes a simulation study in order to explore the performance of the new estimators under normality and contamination. Also, a robust cross validation method for the bandwidth selection is considered. We show an example using real data, in Section 5. Proofs are given in the Appendix.

2 The model and the estimators

2.1 Classical estimators

Assume that we have a sample of n independent variables $(y_i, \mathbf{x}_i^T, t_i)$ in $\mathbb{R}^{p+1} \times M$ with identically distribution to (y, \mathbf{x}^T, t) , where (M, γ) 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^T, t_i) can be represented as

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

where the errors ε_i are independent and independent of (\mathbf{x}_i^T, t_i) and we will assume that ε has symmetric distribution. 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 $1 \leq j \leq p$, then we have that $g(t) = \phi_0(t) - \boldsymbol{\phi}(t)^T \boldsymbol{\beta}$ and hence, $y - \phi_0(t) = (\mathbf{x} - \boldsymbol{\phi}(t))^T \boldsymbol{\beta} + \varepsilon$. The classical least square estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{ls}$ can be obtained by minimizing

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

with $\hat{\phi}_{0,ls}$ and $\hat{\boldsymbol{\phi}}_{ls}$ are nonparametric kernel estimators of ϕ_0 and $\boldsymbol{\phi}(t)$, respectively. More precisely, the nonparametric estimators $\hat{\phi}_{0,ls}$ and $\hat{\phi}_{j,ls}$ of ϕ_0 and ϕ_j can be defined as (see [21]),

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

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 $K : \mathbb{R} \rightarrow \mathbb{R}$ a non-negative function, d_γ the distance induced by the metric γ , $\theta_t(s)$ the volume density function on (M, γ) and the bandwidth h_n is a sequence of real positive numbers such that $\lim_{n \rightarrow \infty} h_n = 0$ and h_n are smaller than the injectivity radius of (M, γ) ($inj_\gamma M$). As in [13] we consider (M, γ) a d -dimensional compact oriented Riemannian manifold without boundary. Note that in this case $inj_\gamma M > 0$. For a rigorous definition of the volume density function and the injectivity radius see [1] or [14].

The final least square estimator of g can be taken as $\hat{g}_{ls}(t) = \hat{\phi}_{0,ls}(t) - \hat{\boldsymbol{\phi}}_{ls}(t)^T \hat{\boldsymbol{\beta}}_{ls}$. The properties of these estimators have been studied in [11] and in the case of Euclidean data have been widely studied in the literature, see for example [10], [22], [9] and [18].

2.2 Robust estimates

As in the Euclidean setting, the estimators introduced by [21] are a weighted average of the response variables, these estimates are very sensitive to large fluctuations of the variables and so, the final estimator of β can be seriously affected by anomalous data, as mentioned in the Introduction. To overcome this problem, [13] 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 [21]. The second generalizes to our setting the proposal given by [5], who considered robust nonparametric estimates using nearest neighbor weights when the predictors t are on \mathbb{R}^d .

We proposed a class of estimates based on a three-step robust procedure under the partly linear model when some of the predictors takes values on a Riemannian manifolds. The three-step robust estimators are defined as follows:

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

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

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

Note that in **Step 1**, the regression functions correspond to predictors taking values in a Riemannian manifold. Local M -type estimates $\hat{\phi}_{0,R}$ and $\hat{\phi}_{j,R}$ are defined in [13] as the solution of

$$\sum_{i=1}^n w_{n,h}(t, t_i) \Psi \left(\frac{y_i - \hat{\phi}_{0,R}(t)}{\sigma_{0,n}(t)} \right) = 0 \quad \text{and} \quad \sum_{i=1}^n w_{n,h}(t, t_i) \Psi \left(\frac{x_{ij} - \hat{\phi}_{j,R}(t)}{\sigma_{j,n}(t)} \right) = 0 \quad (4)$$

respectively, where the score function Ψ is strictly increasing, bounded and continuous and $\sigma_{0,n}(\tau)$ and $\sigma_{j,n}(\tau)$ $1 \leq j \leq p$ are local robust 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 $\sigma_{j,n}(\tau)$ $1 \leq j \leq p$ can be taken as the local median of the absolute deviations from the local median (local MAD), *i.e.* the MAD (see [15]) with respect to the distributions

$$F_n(y|t = \tau) = \sum_{i=1}^n w_{n,h}(\tau, t_i) I_{(-\infty, y]}(y_i) \quad \text{and} \quad F_{j,n}(x|t = \tau) = \sum_{i=1}^n w_{n,h}(\tau, t_i) I_{(-\infty, x]}(x_{ij}). \quad (5)$$

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 ([15]) and GM-estimators ([19]). On the other hand, high breakdown point estimates with high efficiency as MM-estimates could be evaluated ([26] and [27]).

We consider $\widehat{\beta}_R$ the solution of

$$\sum_{i=1}^n \psi_1 \left((\widehat{r}_i - \widehat{\eta}_i^T \widehat{\beta}_R) / s_n \right) w_1 (\|\widehat{\eta}_i\|) \widehat{\eta}_i = 0, \quad (6)$$

with s_n a robust consistent estimate of σ_ε , $\widehat{r}_i = y_i - \widehat{\phi}_{0,R}(t_i)$, $\widehat{\eta}_i = \mathbf{x}_i - \widehat{\phi}_R(t_i)$, 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 [[20], Chapter 2].

The estimator defined by [21] corresponds to the choice $\Psi(u) = u$ with the estimators of the conditional distribution based on kernel weights defined in (5). Therefore, if we considered the least square estimators of β in Step 2, we obtain the classical estimators proposed in [11]. On the other hand, when (M, γ) is \mathbb{R}^d endowed with the canonical metric, the estimation procedure reduces to proposal introduced in [2]. Details over the procedure to computing the robust nonparametric estimators in Step 1 can be found in [13].

3 Asymptotic results

The theorems of this Section study the asymptotic behavior of the regression parameter estimator of the model 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 \mathbb{R} . As in [21], 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 $1 \leq j \leq n$ be the MAD of the conditional distribution of $y_1|t = \tau$ for $j = 0$ and $x_{1j}|t = \tau$ for $1 \leq j \leq n$, respectively.

3.1 Consistency

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

H1. $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ 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

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

for the functions $G_s(z)$ equal to $F(z|t = s)$ and $F_j(z|t = s)$ 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 $z \in \mathbb{R}$, for the functions $G_s(z)$ equal to $F(z|t = s)$ and $F_j(z|t = s)$ for $1 \leq j \leq p$.

H6. $K : \mathbb{R} \rightarrow \mathbb{R}$ is a bounded nonnegative Lipschitz function of order one, with compact support $[0, 1]$ satisfying $\int_{\mathbb{R}^d} \mathbf{u} K(\|\mathbf{u}\|) d\mathbf{u} = \mathbf{0}$ and $0 < \int_{\mathbb{R}^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|) d\mathbf{u} < \infty$.

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

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

Remark 3.1.1. Assumption H1 is a standard condition in a robustness framework. 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. Assumption H6 is a standard assumption when dealing with kernel estimators. 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.

Theorem 3.1.1. Under the hypothesis H1 to H8, we have that

- a) $|\hat{\beta}_R - \beta| \xrightarrow{a.s.} 0$.
- b) $\sup_{\tau \in M_0} |\hat{g}_R(\tau) - g(\tau)| \xrightarrow{a.s.} 0$.

3.2 Asymptotic distribution

In this Section, we assume that in the **Step 2** of the estimation procedure, the choice for $\hat{\beta}_R$ is given as in (6). 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 $\hat{\beta}_R$ defined as a solution of

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

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

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 $A = E(\psi_1'(\varepsilon/\sigma_\varepsilon)w_1(\|\boldsymbol{\eta}_1\|) \boldsymbol{\eta}_1 \boldsymbol{\eta}_1^T)$ 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 $0 \leq j \leq p$ are continuous with ϕ_j' continuous in M .
- A5. $\widehat{\phi}_j(t)$ the estimates of $\phi_j(t)$ for $0 \leq j \leq p$ have first continuous derivatives in M and

$$n^{1/4} \sup_{t \in M} |\widehat{\phi}_j(t) - \phi_j(t)| \xrightarrow{p} 0, \text{ for } 0 \leq j \leq p, \quad (7)$$

$$\sup_{t \in M} |\nabla \widehat{\phi}_j(t) - \nabla \phi_j(t)| \xrightarrow{p} 0, \text{ for } 0 \leq j \leq p. \quad (8)$$

where $\nabla \xi$ corresponds to the gradient of ξ with $\xi \in \mathcal{F}(M)$ and $\mathcal{F}(M)$ the class of functions $\{\xi \in \mathcal{C}^1(M) : \|\xi\|_\infty \leq 1 \ \|\nabla \xi\|_\infty \leq 1\}$.

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

Theorem 3.2.1. Under the assumptions A1 to A6 we have that

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

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

Remark 3.2.1. To proof the previous result, we will need an inequality for the covering number of $\mathcal{F}(M)$. The Appendix include some results related to the covering number on a Riemannian manifold.

4 Simulation study

In this section, we consider a simulation study designed to evaluate the performance of the robust procedure introduced in Section 2. The main objective of this study is to compare the behavior of the classical and robust estimators under normal samples and contamination. We consider the cylinder endowed with the metric induced by the canonical metric of \mathbb{R}^3 . Because of the computational burden 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 $K(t) = (15/16)(1-t^2)^2 I(|x| < 1)$. In the next subsection, we will describe a robust cross validation procedure that we will use in order to obtain the robust estimates and to compute the classical estimators we consider a classical cross validation described in [11]. The distance d_γ and the volume density function for the cylinder were computed in [14] and [13]. We considered the following model:

The variables (y_i, x_i, t_i) for $1 \leq i \leq n$ were 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 follow a uniform distribution in $(0, 2\pi)$ and the variables s_i are uniform in $(0, 1)$, i.e. t_i have support in the cylinder with radius 1 and height between $(0, 1)$.

The non contaminated cases that denoted with C_0 corresponds to the errors ε_i and η_i are i.i.d. normal with mean 0 and standard deviation 1 and 0.05, respectively. Besides, the so-called contaminations C_1 and C_2 , which correspond to selecting a distribution in a neighborhood of the central normal distribution, 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. The contamination C_1 corresponds to inflating the errors and thus, will affect the variance of the regression estimates.

4.1 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 [17], [28], [6], [8] and [16]. Under a nonparametric regression model with carriers in an Euclidean space for spline-based estimators, [8] introduced a robust cross-validation criterion to select the bandwidth parameter. Robust cross-validation selectors for kernel M-smoothers were considered in [17], [28] and [16], under a fully nonparametric regression model. In the Euclidean setting, for partly linear model, a robust cross-validation criterion in the cases of autoregression models was considered in [3], while a robust plug-in procedure was studied in [7]. When the variables belong in a Riemannian manifold, a robust cross validation procedure was discussed in [13] under a fully nonparametric regression model, while a classical cross-validation procedure under a partly linear models was considered in [11].

We included a robust cross-validation method for the choice of the bandwidth in the case of partly linear models that robustified the proposal given in [11] and generalized the procedure given in [7]. The robust cross-validation method constructs an asymptotically optimal data-driven bandwidth, and thus adaptive data-driven estimators, by minimizing

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

with $\hat{\varepsilon}_i(h) = y_i - \hat{\phi}_{0,-i,h}(t_i) - (\mathbf{x}_i - \hat{\phi}_{-i,h}(t_i))^T \tilde{\beta}$; $\hat{\phi}_{0,-i,h}(t)$ and $\hat{\phi}_{-i,h}(t) = (\hat{\phi}_{1,-i,h}(t), \dots, \hat{\phi}_{p,-i,h}(t))$ denote the robust nonparametric estimators computed with bandwidth h using all the data expect the i -th observation and $\tilde{\beta}$ estimate the regression parameter by applying a robust regression estimate to the residuals $y_i - \hat{\phi}_{0,-i,h}(t_i)$ and $\mathbf{x}_i - \hat{\phi}_{-i,h}(t_i)$ and μ_n and σ_n^2 denote robust estimators of location and scale, respectively.

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

4.2 Simulation results

Table 4.1 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 and R the classical and robust estimators, respectively. Figure 4.1 shows the boxplot of the regression parameter.

	$\text{mean}(\widehat{\beta}_{ls})$	$\text{sd}(\widehat{\beta}_{ls})$	$\text{MSE}(\widehat{\beta}_{ls})$	$\text{MSE}(\widehat{g}_{ls})$
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
	$\text{mean}(\widehat{\beta}_R)$	$\text{sd}(\widehat{\beta}_R)$	$\text{MSE}(\widehat{\beta}_R)$	$\text{MSE}(\widehat{g}_R)$
C_0	2.0646	0.1433	0.0274	0.2431
C_1	2.0198	0.2303	0.0534	0.4897
C_2	2.0109	0.2540	0.0646	1.3580

Table 4.1: Performance of regression parameter and the regression functions under the different contaminations.

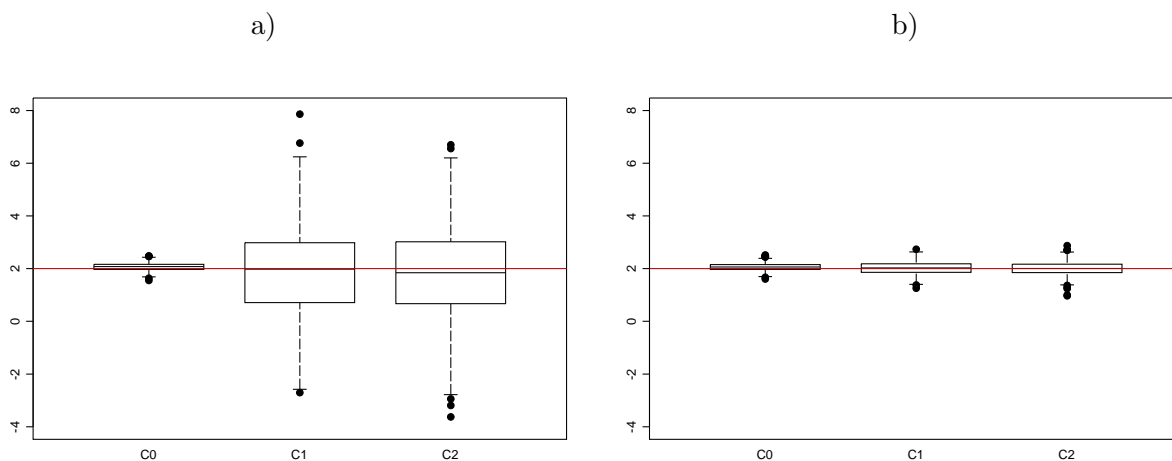


Figure 4.1: Boxplot of a) $\widehat{\beta}_{ls}$ the classical estimators and b) $\widehat{\beta}_R$ the robust estimators under the different contaminations.

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 work showing only a small lack of efficiency under normality. In both cases, the results obtained with the classical estimators are not reliable giving high mean square errors that those corresponding to the robust procedure, under C_1 and C_2 , respectively. This extreme behavior of the classical estimators show its inadequacy when one suspects that the sample can contain outliers.

5 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 [4].

As we comment 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 in <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 measure with the point zero in the north direction and the wind's speed was measure in meter per second. The data was measure daily in the automatic meteorologic station of Pamplona-Larrabide GN, in Navarra, Spain during the year 2004. In our study, we consider a random sample of this dataset.

In Figure 5.1, we can see that the humidity and the insolation follows a lineal relation less in the points contained in the ellipse. 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 nonparametrically, belong to a cylinder. In the smoothing procedure, we consider the quadratic kernel $K(t) = (15/16)(1 - t^2)^2 I(|x| < 1)$ and we select the bandwidth using the robust cross validation procedure for the robust estimators and the classical cross validation described in [11] for the classical estimators.

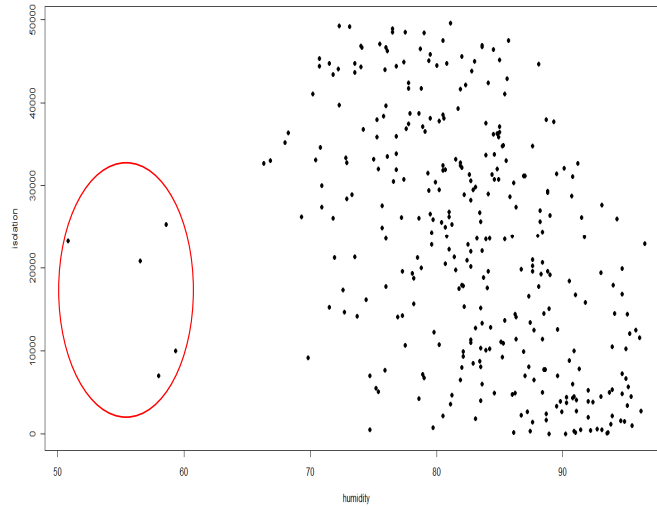


Figure 5.1: Scatter plot between the insolation and humidity. The dots in the ellipse correspond to the potential outliers.

In a first step, we apply the classical and robust methods to obtain an estimator of the regression parameter using all the data. The results were $\hat{\beta}_{ls} = -1032.869$ and $\hat{\beta}_R = -1246.856$. Also, based in the asymptotic results obtained in Theorem 3.2.1, we calculated a confidence interval with level 0.05 in each case. To computed these confidence intervals, we estimated the unknown quantities. The result of the classical confidence interval was $CCI_{0.05}(\beta) = (-1229.9451 - 835.7935)$ and the confidence interval based in the robust estimation was $RCI_{0.05}(\beta) = (-1453.658, -1040.053)$. On the other hand, we calculated the classical estimator using the data except the potential outliers, the result was $\hat{\beta}_{ls} = -1294.620$ and its confidence interval was $CCI_{0.05}(\beta) = (-1502.983, -1086.257)$. Thus, if we estimate the regression parameter with the classical approach when the dataset have outliers, the conclusion can be different. For example. in the classical case with all the data, the hypothesis that $\beta = -1000$ is rejected, while the conclusions with the classical estimator without the outliers or the robust estimators with all the data does not reject the null hypothesis.

Acknowledgments

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.

A Appendix

A.1 Proof of Theorem 3.1.1.

a) Denote by $\hat{r}_i = y_i - \hat{\phi}_{0,R}(t_i)$, $\hat{\boldsymbol{\eta}}_i = \mathbf{x}_i - \hat{\boldsymbol{\phi}}_R(t_i)$, $\boldsymbol{\eta}_i = \mathbf{x}_i - \boldsymbol{\phi}(t_i)$ and $r_i = y_i - \phi_0(t_i)$. We note that $r_i = \boldsymbol{\eta}_i^\top \boldsymbol{\beta} + \varepsilon_i$ and let $\widehat{P}_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(\hat{r}_i, \hat{\boldsymbol{\eta}}_i)$. It is well known that the robust regression estimates can be written as a functional of the empirical distribution. More precisely, $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\widehat{P}_n)$ where $\boldsymbol{\beta}(\cdot)$ is continuous at P , the common distribution of $(r_i, \boldsymbol{\eta}_i^\top)^\top$. Therefore, 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 $f : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ we have that $|E_{\widehat{P}_n} f - E_P f| \xrightarrow{a.s.} 0$.

Note that

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

where $C_1 \subset \mathbb{R}^{p+1}$ and $M_0 \subset M$ are compact sets such that for any $\varepsilon > 0$ $P(C) > 1 - \varepsilon / (4\|f\|_\infty)$ with $C = C_1 \times M_0$.

Under the assumptions by Theorem 3.3 of [13], we have that

$$\sup_{t \in M_0} |\hat{\phi}_{j,R}(t) - \hat{\phi}_j(t)| \xrightarrow{a.s.} 0$$

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 that for any $\omega \notin \aleph$ we obtain that

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

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 $\max_{1 \leq j \leq p+1} |u_j - v_j|, u, v \in \bar{C}_1$ entails $|f(u) - f(v)| \leq \frac{\varepsilon}{2}$. Thus, we have that for $\omega \notin \aleph$ and n large enough $\max_{0 \leq j \leq p} \sup_{t \in M_0} |\hat{\phi}_j(t) - \phi_j(t)| < \delta$ and then, for $1 \leq i \leq n$, we obtain that

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

that conclude the proof.

b) \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 [23], [24] and [25]. In particular, [25] obtained an upper-bound to the covering number for $\mathcal{F}(M)$ when M is a bounded, convex subset of \mathbb{R}^d . For the convenience of the reader, we have included the following remark (see [12]).

Remark A.1. Let $N(\delta)$ be the minimal number of balls with radius δ needed to cover (M, δ) . A δ -filling is a maximal family of pairwise disjoint open balls of radius δ . We denote by $D(\delta)$, the packing number, *i.e.* the maximum number of such balls. It is easy to see that $N(2\delta) \leq D(\delta)$. Let $diam_{(M, \gamma)}$ be the diameter of (M, γ) and consider $\kappa \in \mathbb{R}$ 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 scalar curvature $2(d-1)\kappa$ then the inequality is attained. Note that if $\kappa > 0$ since Myers's Theorem [12], (M, γ) is necessarily a compact manifold 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 [12]) 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,

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

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

Lemma A.1. Let $\mathcal{F}(M) = \{\xi \in C^1(M) : \|\xi\|_\infty \leq 1, \|\nabla \xi\|_\infty \leq 1\}$, 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 $\mathcal{A} = \{B(a_1, \delta), \dots, B(a_N, \delta)\}$ 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 $\tilde{\xi} = \sum_{i=1}^N \delta \left[\frac{\xi(a_i)}{\delta} \right] I_{D_i}$ 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 $|\tilde{\xi}(x) - \xi(x)| \leq |\tilde{\xi}(x) - \xi(a_k)| + |\xi(a_k) - \xi(x)|$. Since $\tilde{\xi}(a_k) = \tilde{\xi}(x)$ and $\xi(a_k) = \tilde{\xi}(a_k) + \delta(\frac{\xi(a_k)}{\delta} - [\frac{\xi(a_k)}{\delta}]) = \tilde{\xi}(a_k) + \delta B$ with $0 \leq B < 1$, we have that $|\tilde{\xi}(x) - \xi(x)| \leq 2\delta$.

For the first value $\tilde{\xi}(a_1)$ of a generic function $\tilde{\xi}$, we have $[\frac{1}{\delta}] + 1$ possibilities. Since,

$$|\tilde{\xi}(a_k) - \tilde{\xi}(a_{k-1})| \leq |\tilde{\xi}(a_k) - \xi(a_k)| + |\xi(a_k) - \xi(a_{k-1})| + |\xi(a_{k-1}) - \tilde{\xi}(a_{k-1})| \leq 4\delta.$$

Therefore, for each value of $\tilde{\xi}(a_{k-1})$ we can choose 9 possibilities for $\tilde{\xi}(a_k)$. Then it is easy to verify that

$$N(2\delta, \mathcal{F}(M), \|\cdot\|_\infty) \leq \left([\frac{1}{\delta}] + 1\right) 9^N.$$

which finish the proof. \square

Remark A.2. Since $N(\delta, \mathcal{F}(M), L^2(Q)) \leq N(\delta, \mathcal{F}(M), \|\cdot\|_\infty)$ then 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 3.2.1.

Using a Taylor expansion around $\hat{\beta}_R$ we have that $S_n = A_n(\hat{\beta}_R - \beta)$ where

$$\begin{aligned} S_n &= \frac{1}{n} \sum_{i=1}^n \psi_1 \left((\hat{r}_i - \hat{\boldsymbol{\eta}}_i^\top \beta) / s_n \right) w_1(\|\hat{\boldsymbol{\eta}}_i\|) \hat{\boldsymbol{\eta}}_i \\ A_n &= \frac{1}{n} \sum_{i=1}^n \psi_1' \left((\hat{r}_i - \hat{\boldsymbol{\eta}}_i^\top \tilde{\beta}) / s_n \right) w_1(\|\hat{\boldsymbol{\eta}}_i\|) \hat{\boldsymbol{\eta}}_i \hat{\boldsymbol{\eta}}_i^\top. \end{aligned}$$

where $\tilde{\beta}$ is an intermediate point between β and $\hat{\beta}_R$. Analogous arguments to those used in Lemma 2 in [2] 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

$$\sqrt{n} \left[S_n - \frac{1}{n} \sum_{i=1}^n \psi_1(\varepsilon_i / s_n) w_1(\|\boldsymbol{\eta}_i\|) \boldsymbol{\eta}_i \right] \xrightarrow{p} 0, \quad (9)$$

$$\sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n \psi_1(\varepsilon_i / s_n) w_1(\|\boldsymbol{\eta}_i\|) \boldsymbol{\eta}_i - \frac{1}{n} \sum_{i=1}^n \psi_1(\varepsilon_i / \sigma_\varepsilon) w_1(\|\boldsymbol{\eta}_i\|) \boldsymbol{\eta}_i \right] \xrightarrow{p} 0. \quad (10)$$

We first prove (9). Using a Taylor expansion of order two, we have that the following decomposition.

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

where

$$\begin{aligned} S_{n1} &= \frac{\sqrt{n}}{n} \sum_{i=1}^n \psi_1'(\varepsilon_i / s_n) [\hat{\boldsymbol{\gamma}}^\top(t_i) \beta - \hat{\gamma}_0(t_i)] w_1(\|\boldsymbol{\eta}_i\|) \boldsymbol{\eta}_i \\ S_{n2} &= \frac{s_n \sqrt{n}}{n} \sum_{i=1}^n \psi_1(\varepsilon_i / s_n) [w_1(\|\hat{\boldsymbol{\eta}}_i\|) \hat{\boldsymbol{\eta}}_i - w_1(\|\boldsymbol{\eta}_i\|) \boldsymbol{\eta}_i] \\ S_{n3} &= \frac{s_n \sqrt{n}}{n} \sum_{i=1}^n [\psi_1(\hat{r}_i - \hat{\boldsymbol{\eta}}_i^\top \beta / s_n) - \psi_1(\varepsilon_i / s_n)] w_1(\|\hat{\boldsymbol{\eta}}_i\|) [\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i] \\ S_{n4} &= \frac{\sqrt{n}}{2n} \sum_{i=1}^n \psi_1''(\varepsilon_i / s_n) [\hat{\boldsymbol{\gamma}}^\top(t_i) \beta - \hat{\gamma}_0(t_i)]^2 w_1(\|\hat{\boldsymbol{\eta}}_i\|) \boldsymbol{\eta}_i \\ S_{n5} &= \frac{\sqrt{n}}{n} \sum_{i=1}^n \psi_1(\varepsilon_i / s_n) [\hat{\boldsymbol{\gamma}}^\top(t_i) \beta - \hat{\gamma}_0(t_i)] [w_1(\|\hat{\boldsymbol{\eta}}_i\|) - w_1(\|\boldsymbol{\eta}_i\|)] \boldsymbol{\eta}_i \end{aligned}$$

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

Let

$$\begin{aligned}\mathcal{J}_{1n}^{(j)}(\sigma, \xi) &= \frac{\sqrt{n}}{n} \sum_{i=1}^n f_{1,\sigma,\xi}^{(j)}(r_i, \boldsymbol{\eta}_i, t_i) \\ &= \frac{\sqrt{n}}{n} \sum_{i=1}^n \psi_1' \left(\frac{r_i - \boldsymbol{\eta}_i^T \boldsymbol{\beta}}{\sigma} \right) \xi(t_i) w_1(\|\boldsymbol{\eta}_i\|) (\boldsymbol{\eta}_i)_j \\ \mathcal{J}_{2n}^{(j)}(\sigma, \boldsymbol{\xi}) &= \frac{\sqrt{n}}{n} \sum_{i=1}^n f_{2,\sigma,\boldsymbol{\xi}}^{(j)}(r_i, \boldsymbol{\eta}_i, t_i) \\ &= \frac{\sigma \sqrt{n}}{n} \sum_{i=1}^n \psi_1 \left(\frac{r_i - \boldsymbol{\eta}_i^T \boldsymbol{\beta}}{\sigma} \right) [w_1(\|\boldsymbol{\eta}_i + \boldsymbol{\xi}\|) (\boldsymbol{\eta}_i + \boldsymbol{\xi}(t_i))_j - w_1(\|\boldsymbol{\eta}_i\|) (\boldsymbol{\eta}_i)_j]\end{aligned}$$

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

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

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

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 [2], it is easy to see that the covering number of the classes \mathcal{F}_1 and \mathcal{F}_2 satisfy

$$\begin{aligned}N(C_1 \varepsilon, \mathcal{F}_1, L^2(Q)) &\leq N(\varepsilon, \mathcal{F}(M), L^2(Q)) N(\varepsilon, (\sigma_\varepsilon/2, 2\sigma_\varepsilon), |\cdot|) \\ N(C_2 \varepsilon, \mathcal{F}_2, L^2(Q)) &\leq N^p(\varepsilon, \mathcal{F}(M), L^2(Q)) N(\varepsilon, (\sigma_\varepsilon/2, 2\sigma_\varepsilon), |\cdot|)\end{aligned}$$

where Q is any probability measure. Since Remark A.2, the covering number of $\mathcal{F}(M)$ satisfies that $\log N(\varepsilon, \mathcal{F}(M), L^2(Q)) < A\varepsilon^{-d}$. Therefore, we get that these classes 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,

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

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

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

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

$$\begin{aligned} &\leq \frac{1}{\epsilon} E \left(\sup_{f \in \mathcal{F}_{1,\delta}} \left| \frac{\sqrt{n}}{n} \sum_{i=1}^n f(r_i, \boldsymbol{\eta}_i, t_i) \right| \right) + \delta \\ &\leq \frac{1}{\epsilon} \mathcal{G}(\delta, \mathcal{F}_1) + \delta \end{aligned}$$

where $\mathcal{G}(\delta, \mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L^2(Q))} d\epsilon$, then the fact that \mathcal{F}_1 satisfies the uniform-entropy conditions we get that $\lim_{\delta \rightarrow 0} \mathcal{G}(\delta, \mathcal{F}_1) = 0$, therefore $S_{1n} \xrightarrow{p} 0$. Similarly for $\mathcal{J}_{2n}(s_n, \hat{\gamma})$ and the class \mathcal{F}_2 and we get that $S_{2n} \xrightarrow{p} 0$.

The proof of (10), follows using analogous arguments that those considered in (9). \square

References

- [1] Besse, A. (1978). Manifolds all of whose Geodesics are Closed. *Springer-Verlag*.
- [2] Bianco, A. and Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Plann. Inference.* **122**, 229–252.
- [3] Bianco, A. Boente, G. (207). Robust estimators under semi-parametric partly linear autoregression: asymptotic behaviour and bandwidth selection. *J. Time Ser. Anal.* **28**, 274–306.
- [4] Bird, R. and Hulstrom, R. (1980). Direct insolation models. Solar Energy Research Institute.
- [5] Boente, G. and Fraiman, R. (1989). Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, **29**, 180-198.
- [6] Boente, G.; Fraiman, R. and Meloche, J. (1997). Robust plug-in bandwidth estimators in non-parametric regression. *J. Statist. Plann. Inference.* **57**, 109142.
- [7] Boente, G. and Rodriguez, D. (2008). Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis. *Computational Statistics & Data Analysis.* **52**, 2808-2828.
- [8] Cantoni, E. and Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statist. Comput.* **11**, 141146.
- [9] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16**, 136-146.
- [10] 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.
- [11] Gonzalez-Manteiga, W.; Henry, G. and Rodriguez, D. (2012). Partly linear models on Riemannian manifolds. *Journal of Applied Statistics.* **39**, 8, 1797-1809.
- [12] Gallot, S.; Hulin, D. and Lafontaine, J. (1986). Riemannian Geometry, 3th edition. *Springer*.
- [13] Henry, G. and Rodriguez, D. (2009). Robust Nonparametric Regression on Riemannian Manifolds. *Journal of Nonparametric Statistics.* **21**, 5, 611-628.
- [14] Henry, G. and Rodriguez, D. (2009). Kernel Density Estimation on Riemannian Manifolds: Asymptotic Results. *Journal Math. Imaging Vis.* **43**, 235-639.

- [15] Huber, P. (1981). *Robust Statistics*, Wiley, New York.
- [16] Leung, D. (2005). Cross-validation in nonparametric regression with outliers. *Ann. Statist.* **33**, 2291-2310.
- [17] Leung, D.; Marriott, F. and Wu, E. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics.* **4**, 333-339.
- [18] Liang, H. (2000). Asymptotic of nonparametric part in partly linear models with measurement error in the nonparametric part. *Journal of statistical planning and inference.* **86**, 51-62.
- [19] Mallows, C. (1975). On some topics in robustness. Technical Memorandum, AT&T Bell Laboratories, Murray Hill.
- [20] Maronna, R., Martin, D. and Yohai, V. (2006), *Robust Statistics: Theory and Methods*, New York: Wiley.
- [21] Pelletier, B. (2006). Nonparametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics.* **18**, 57-67.
- [22] Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B.* **50**, 413-436.
- [23] van der Geer, Sara. (2000). *Empirical Processes in M-Estimation.* Cambridge University Press.
- [24] Van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- [25] Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics.* New York: Springer.
- [26] Yohai, V. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15**, 642-656.
- [27] Yohai, V. and Zamar, R. (1988). High breakdown estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83**, 406-413.
- [28] Wang, F. and Scott, D. (1994). The L1 method for robust nonparametric regression. *J. Amer. Statist. Assoc.* **89**, 6576.