Statistics and Probability Letters 134 (2018) 128–133

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Local linear estimate of the nonparametric robust regression in functional data

Faiza Belarbi [a], Souheyla Chemikh [a], Ali Laksaci [b,*]

[a] Laboratoire de statistique et processus stochastiques, Univ. Djillali Liabès, Algeria
[b] Department of Mathematics, College of Science, King Khalid University, P.O. Box: 9004, Postal Code: 61413, Abha, Saudi Arabia

ARTICLE INFO

Article history:
Received 6 March 2017
Received in revised form 26 October 2017
Accepted 7 November 2017
Available online 22 November 2017

Keywords:
Functional data analysis
Local linear method
Robust estimation
Asymptotic normality
Almost complete convergence

ABSTRACT

In this paper, we study the robust estimation of the functional local linear regression model. The main results of this work are the establishment of the almost complete convergence as well as the asymptotic normality for the constructed estimator.

1. Introduction

The nonparametric functional data analysis is an emergent field of research in nonparametric statistics. In this context, the modelization of the relationship between a scalar response variable and functional covariate has attracted a considerable interest. The most used approaches are based on local constant fits. In this work, we propose to model this relationship by using the local linear M-regression.

It is well known that the local linear method has several advantages over the local constant fitting. In particular, it allows to reduce the bias term in a various situations. We refer to Fan and Gijbels (1996) for more discussions on the importance of this approach. Noting that the robust local linear estimation has investigated by many authors in the multivariate case. See Fan et al. (1994), for the i.i.d. case, Cai and Ould Said (2003) for the α-mixing case. We return to Boente et al. (2009) for some asymptotic results in the robust local constant method for both procedures (kernel and KNN methods). In this work, we will focus on the case where the covariate is of infinite dimension. It should be noted that, these questions of statistical analysis of infinite dimensional data are arising more and more in applied statistics. For a bibliographical survey or an overview on recent developments in this topic, we cite, for example, Geenens (2011), Horvàth and Kokoszka (2012), Zhang (2014), Hsing and Eubank (2015) and Goia and Vieu (2016).

The local linear method has been considered for functional data. The first interesting results, on this topic, were obtained by Baìllo and Grané (2009). They established the L2 consistency of a local linear estimator of the regression operator when the explanatory variable is Hilbertian. Barrientos-Marin et al. (2010) consider an alternative fast version of the functional local linear estimator which can be used for more general functional regressor. This last contribution gives the almost complete convergence (with rate) of the proposed estimate. Berlinet et al. (2011) propose another local linear estimate based on the inverse of the local covariance operator of the functional explanatory variable. We return to Zhou and Lin (2016) for the

* Corresponding author.
E-mail addresses: faiza_belarbi@yahoo.fr (F. Belarbi), chemikh_souheyla@yahoo.com (S. Chemikh), alilak@yahoo.fr (A. Laksaci).

https://doi.org/10.1016/j.spl.2017.11.003
0167-7152/

asymptotic normality of the functional local linear regression estimate. For a more recent advances on the functional kernel estimate or on its alternative functional smoothers that is the KNN-method we cite Kara-Zaitri et al. (2017a, b). All these studies focus mainly on the classical regression. However, it is well known that this model is very sensitive to outliers and do not perform well when the errors are heavy-tailed. Such kind of data is observed very often in econometrics and finance as well as in many other applied fields. To attenuate the lack of robustness of this model, we propose in this work to robustify the functional local linear regression model.

In recent years, there is a growing body of studies that consider the nonparametric robust regression for functional data. We cite, for instance, Chen and Zhang (2009) for previous results and Boente and Vahnovan (2015) for recent advances and references. All these works use an estimation procedure based on the local constant fits. As far as we are aware, there has not been any attempt so far to study the robust-type of local linear regression in functional data. This is the main goal of this paper. Precisely, we combine the ideas Barrientos-Marin et al. (2010) on the functional local linear method with those of the robustness developed by Fan et al. (1994) to construct an estimator resistant to the presence of outliers or heteroscedasticity of data and inherits many nice statistical properties of the local linear approach. The asymptotic behavior of this approach is studied under some standard conditions of nonparametric functional data analysis. Precisely, under the concentration properties on small balls of the probability measure of this functional variable, we establish the almost complete consistency (with rate) and the asymptotic normality of this estimator.

This paper is organized as follows: We present our model in Section 2. The needed conditions and the main results are given in Section 3. In Section 4 we highlight the main features of our approach and we compare it to the existing approaches. We present also in this section some prospects of the present contribution. A sketch of the proof of the main result is done in Appendix. See the details of the proof in Belarbi et al. (2017).

2. The robust local linear estimator

Consider n independent pairs of random variables (X_i, Y_i) for $i = 1, \ldots, n$ that we assume drawn from the pair (X, Y). The latter is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space and d denotes a semi-metric. The object of this paper is to study the co-variation between X_i and Y_i by the nonparametric robust regression function. For $x \in \mathcal{F}$ the nonparametric robust regression, denoted by θ_x, is defined as the unique minimizer of

$$\theta_x = \arg\min_{t\in\mathbb{R}} \mathbb{E}\left[\rho(Y-t)|X=x\right] \tag{1}$$

where $\rho(.)$ is a real-valued Borel function satisfying some regularity conditions to be stated below. This kind of models belongs to the class of M-estimates introduced by Huber (1964). It covers and includes many usual nonparametric models, for example, for $\rho(y) = y^2$ we obtain the classical regression, $\rho(y) = |y||\alpha - \mathbb{I}_{y<0}|$ leads to the αth conditional quantile. The αth conditional expectile is obtained by setting $\rho(y) = y^2|\alpha - \mathbb{I}_{y<0}|$. For more others examples we refer the reader to Stone (2005).

Recall that the basic idea of local linear fitting consists in approximating locally the nonparametric model by a linear function. In functional statistics, there are several ways for extending this approach (see, Baìllo and Grané (2009) or Barrientos-Marin et al. (2010) for some examples). In this paper, we adopt the fast version proposed by Barrientos-Marin et al. (2010) for which the function θ_x is approximated by

$$\forall z \quad \text{in neighborhood of } x \qquad \theta_z = a + b\beta(z,x)$$

where a and b are estimated by $\widehat{a}$ and $\widehat{b}$ are solution of

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho(Y_i - a - b\beta(X_i,x))K\left(h^{-1}\delta(x,X_i)\right) \tag{2}$$

where $\beta(.,.)$ is a known function from $\mathcal{F}\times\mathcal{F}$ into $\mathbb{R}$ such that, $\forall\xi \in \mathcal{F}$, $\beta(\xi,\xi) = 0$, with K is kernel and $h = h_n$ is a sequence of positive real numbers and $\delta(.,.)$ is a function of $\mathcal{F}\times\mathcal{F}$ such that $d(.,.) = |\delta(.,.)|$. It is clear that, under this consideration, we can write

$$\theta_x = a \quad \text{and} \quad \widehat{\theta_x} = \widehat{a}.$$

We point out that unlike to the classical regression case studied by Barrientos-Marin et al. (2010) the robust local linear estimator $\widehat{\theta_x}$ cannot explicitly expressed. Thus, the establishment of the asymptotic proprieties of our estimate is very difficult, it requires some additional tools.

3. Main results

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. Moreover, x denotes a fixed point in $\mathcal{F}$, N_x denotes a fixed neighborhood of x. For $i = 1, \ldots, n$, we denote by $K_i = K(h^{-1}\delta(x, X_i))$, and $\beta_i = \beta(X_i, x)$. Furthermore, we put $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \le \delta(x, X) \le r_1)$ and we assume the following hypotheses: