Computational Statistics and Data Analysis 65 (2013) 46–55

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Robust functional linear regression based on splines

Ricardo A. Maronna [a,*], Victor J. Yohai [b]

[a] *Department of Mathematics, School of Exact Sciences, Universidad Nacional de La Plata, Argentina*

[b] *Department of Mathematics, School of Exact and Natural Sciences, Universidad de Buenos Aires and CONICET, Argentina*

ARTICLE INFO

Article history:
Received 22 February 2011
Received in revised form 16 November 2011
Accepted 16 November 2011
Available online 26 November 2011

Keywords:
MM estimate
Natural splines
Robust ridge estimator

ABSTRACT

Many existing methods for functional regression are based on the minimization of an L_2 norm of the residuals and are therefore sensitive to atypical observations, which may affect the predictive power and/or the smoothness of the resulting estimate. A robust version of a spline-based estimate is presented, which has the form of an MM estimate, where the L_2 loss is replaced by a bounded loss function. The estimate can be computed by a fast iterative algorithm. The proposed approach is compared, with favorable results, to the one based on L_2 and to both classical and robust Partial Least Squares through an example with high-dimensional real data and a simulation study.[1]

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

We consider the analysis of data described by a linear functional regression model. That is, our data are independent identically distributed (i.i.d.) pairs (X_i, y_i), $i = 1, \ldots, n$, where $y_i \in R$ and $X_i(.)$ are random functions defined on an interval I, such that

$$y_i = \alpha_0 + \int_I \alpha(t) X_i(t) dt + e_i, \quad i = 1, \ldots, n, \tag{1}$$

where the number α_0 and the function $\alpha(t)$ are unknown, and $\{e_i\}$ are i.i.d. random errors independent of $\{X_i\}$. In practice one actually observes at given points $t_1 < \cdots < t_p$ in I the values $x_{ij} = X_i(t_j)$. Henceforth we shall denote $\mathbf{X} = [x_{ij}] \in R^{n\times p}$ and $\mathbf{y} = [y_i] \in R^n$;

These data sets are often high-dimensional, in many cases with $p \gg n$. The functional framework allows to profit from qualitative assumptions like smoothness of underlying curves. This type of regression model was first considered in Ramsay and Dalzell (1991). Ramsay and Silverman (2002, 2005) and Ferraty and Vieu (2006), present several case studies demonstrating the advantages of these models. Among recent applications, Goldsmith et al. (2010) present an application to diffusion tensor imaging (DTI) tractography, and Delaigle et al. (2009) deal with a meteorological application. Cardot et al. (2005, 2006) present the theory and applications of quantile regression for functional data.

One of the most important approaches for the estimation of α_0 and α is regularization through a penalized least squares approach after expanding in some basis such as splines: see Ramsay and Dalzell (1991), Eilers and Marx (1996), Marx and Eilers (1999), Cardot et al. (2003). Crambes et al. (2009) proposed a smoothing splines approach prolonging previous work from Cardot et al. (2007). They show that the rates of convergence of their estimators are optimal in the sense

* Correspondence to: Departamento de Matemáticas, Facultad de Ciencias Exactas, C.C. 172, La Plata 1900, Argentina.
E-mail address: rmaronna@retina.ar (R.A. Maronna).

[1] Matlab code for the proposed procedure is provided as supplemental material.

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.11.014

that they are minimax over large classes of distributions of X_i and of functions α. Their approach boils down to an easy to implement procedure. Recently Wang et al. (2012) proposed a spline-based nonparametric transformation model for functional regression.

Most approaches to functional regression are based on minimizing some L_2 norm, and are therefore sensitive to outliers, which calls for the development of robust methods. There are numerous articles on robust methods for functional data. In particular, Crambes et al. (2008) propose a robust estimator for nonparametric models, and Gervini (submitted for publication) deals with robust regression between two stochastic processes; But we are not aware of any robust approach for model (1). The purpose of this article is to propose a robust version of the estimator proposed by Crambes et al. (2009), based on the approach of MM estimation (Yohai, 1987).

Section 2 describes the proposed estimator, the advantages of which are demonstrated in Sections 3 and 4 through their performances with real and simulated data sets, respectively. The computing times of the different estimators are compared in Section 5. Finally Section 6 contains the conclusions of the study.

2. The proposed estimator

We first describe the estimator proposed by Crambes et al. (2009). Let $\widetilde{\mathbf{X}} = [\widetilde{x}_{ij}]$ and $\widetilde{\mathbf{y}} = [\widetilde{y}_i]$ be $\mathbf{X}$ and $\mathbf{y}$ centered by their averages. The estimator is the function $\widehat{\alpha}(t)$ in the Sobolev space $W^{m,2}(I)$ (see e.g. Adams and Fournier, 2003) such that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_i - \frac{1}{p}\sum_{j=1}^{p}\widehat{\alpha}\left(t_j\right)\widetilde{x}_{ij}\right)^2 + \lambda\left(\frac{1}{p}\sum_{j=1}^{p}\pi_{\widehat{\alpha}}\left(t_j\right)^2 + \int_I \widehat{\alpha}^{(m)}(t)^2\,dt\right) = \min \tag{2}$$

where in general $\alpha^{(m)}(t)$ denotes the m-th derivative of $\alpha(t)$, $\lambda > 0$ is a penalty parameter and

$$\pi_{\widehat{\alpha}}(t) = \sum_{l=1}^{m}\gamma_{b,l}t^{l-1} = \arg\min_{\pi}\sum_{j=1}^{p}\left(\widehat{\alpha}\left(t_j\right) - \pi\left(t_j\right)\right)^2,$$

where π ranges over the polynomials in t of degree $m-1$. The $\pi_{\widehat{\alpha}}$ terms ensure the existence of a unique solution. The terms with $\int_I \widehat{\alpha}^{(m)}(t)^2 dt$ penalize the solutions' roughness.

The problem (2) has an explicit solution. Let $\mathbf{b}(t) = \left(b_1(t), \ldots, b_p(t)\right)'$ be a functional basis of the space $NS^m(t_1, \ldots, t_p)$ of natural splines of order $2m$ on I with knots $t_1, \ldots, t_p$. Call $\mathbf{B}$ the $p \times p$ matrix with elements $b_i\left(t_j\right)$. Put $\mathbf{U} = \int_I \mathbf{b}^{(m)}(t)\mathbf{b}^{(m)}(t)'dt$ and let $\mathbf{P}_m$ be the $p \times p$ projection matrix projecting R^p onto the m-dimensional linear space of all (discretized) polynomials of degree $m-1$; i.e., $\mathbf{P}_m = \mathbf{G}\mathbf{G}^+$, where $\mathbf{G}$ has elements $g_{jk} = t_j^k$, $j = 1, \ldots, p$, $k = 0, \ldots, m-1$, and $\mathbf{G}^+$ stands for its pseudo-inverse. Let $\mathbf{A}_m^* = \mathbf{B}^{+\prime}\mathbf{U}\mathbf{B}^+$ and let

$$\mathbf{A}_m = \mathbf{P}_m + p\mathbf{A}_m^*, \tag{3}$$

Then Crambes et al. (2009) show that $\widehat{\alpha} \in NS^m(t_1, \ldots, t_p)$ and that the solution for the vector $\widehat{\alpha} = \left(\widehat{\alpha}(t_1), \ldots, \widehat{\alpha}\left(t_p\right)\right)'$ is

$$\widehat{\alpha} = \arg\min_{\mathbf{a}\in R^p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_i - p^{-1}\widetilde{\mathbf{x}}_i'\mathbf{a}\right)^2 + p^{-1}\lambda\mathbf{a}'\mathbf{A}_m\mathbf{a}\right\}, \tag{4}$$

where $\widetilde{\mathbf{x}}_i'$ is the i-th row of $\widetilde{\mathbf{X}}$.

As a robustification of the former approach, we propose to find a function $\widehat{\alpha}$ and a number $\widehat{\alpha}_0$ such that

$$\widehat{\sigma}_{\text{ini}}^2\sum_{i=1}^{n}\rho\left(\frac{y_i - \widehat{\alpha}_0 - p^{-1}\sum_{j=1}^{p}x_{ij}\widehat{\alpha}\left(t_j\right)}{\widehat{\sigma}_{\text{ini}}}\right) + \lambda\left(\frac{1}{p}\sum_{j=1}^{p}\pi_{\widehat{\alpha}}\left(t_j\right)^2 + \int_I \widehat{\alpha}^{(m)}(t)^2 dt\right) = \min, \tag{5}$$

where ρ is a bounded "ρ-function" in the sense of (Maronna et al., 2006), i.e., $\rho(t)$ is a nondecreasing function of $|t|$ with $\sup_t \rho(t) = 1$; and $\widehat{\sigma}_{\text{ini}}$ is a residual M-estimator of scale from an initial estimator (to be described later). The factor $\widehat{\sigma}_{\text{ini}}^2$ before the summation is employed to make the estimator coincide with the classic one when $\rho(t) = t^2$.

It is not difficult to show that again $\widehat{\alpha} \in NS^m(t_1, \ldots, t_p)$, since $\{\pi_{\widehat{\alpha}}\left(t_j\right), j = 1, \ldots, p\}$ depends only on the values of $\widehat{\alpha}$ at $t_1, \ldots, t_p$, and it is well-known that given these values, the function $\widehat{\alpha}$ minimizing the integral in (5) belongs to $NS^m(t_1, \ldots, t_p)$. Let $\widehat{\alpha}_1 = \left(\widehat{\alpha}(t_1), \ldots, \widehat{\alpha}\left(t_p\right)\right)'$ and $\mathbf{x}_i = \left(x_{i1}, \ldots, x_{ip}\right)'$. Then it follows that (5) implies

$$\widehat{\sigma}_{\text{ini}}^2\sum_{i=1}^{n}\rho\left(\frac{y_i - \widehat{\alpha}_0 - p^{-1}\mathbf{x}_i'\widehat{\alpha}_1}{\widehat{\sigma}_{\text{ini}}}\right) + p^{-1}\lambda\widehat{\alpha}_1'\mathbf{A}_m\widehat{\alpha}_1 = \min. \tag{6}$$

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات