



Functional linear regression after spline transformation

Guochang Wang^a, Nan Lin^b, Baoxue Zhang^{a,*}

^a Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun, China

^b Department of Mathematics, Washington University in St. Louis, One Brookings Drive, Saint Louis, MO 63130, USA

ARTICLE INFO

Article history:

Received 4 March 2011
Received in revised form 18 June 2011
Accepted 6 September 2011
Available online 1 October 2011

Keywords:

Functional linear regression
Transformation
B-spline
Model selection
Mixed-data canonical correlation analysis

ABSTRACT

Functional linear regression has been widely used to model the relationship between a scalar response and functional predictors. If the original data do not satisfy the linear assumption, an intuitive solution is to perform some transformation such that transformed data will be linearly related. The problem of finding such transformations has been rather neglected in the development of functional data analysis tools. In this paper, we consider transformation on the response variable in functional linear regression and propose a nonparametric transformation model in which we use spline functions to construct the transformation function. The functional regression coefficients are then estimated by an innovative procedure called mixed data canonical correlation analysis (MDCCA). MDCCA is analogous to the canonical correlation analysis between two multivariate samples, but is between a multivariate sample and a set of functional data. Here, we apply the MDCCA to the projection of the transformation function on the B-spline space and the functional predictors. We then show that our estimates agree with the regularized functional least squares estimate for the transformation model subject to a scale multiplication. The dimension of the space of spline transformations can be determined by a model selection principle. Typically, a very small number of B-spline knots is needed. Real and simulation data examples are further presented to demonstrate the value of this approach.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Functional data, such as data in the form of curves and images, are becoming more popularly seen. There has been a rapidly increasing level of interest in developing statistical tools for functional data analysis (FDA). An extensive review of FDA can be found in Ramsay and Silverman (2002, 2005), Ferraty and Vieu (2006) and Ferraty and Romain (2010). Many authors have made an effort to adapt existing multivariate regression methods to the functional regression model with a scalar response

$$y = r(x) + \varepsilon, \quad (1)$$

where x is a random variable valued in $L^2([a, b])$, the space of square integrable functions from $[a, b]$ into \mathbb{R} , y and ε are real scalar variables, $r : L^2([a, b]) \rightarrow \mathbb{R}$ is a regression operator, and the inner product in $L^2([a, b])$ is defined by $\langle f, g \rangle = \int_a^b f(t)g(t)dt$. The regression operator r in this model can be modeled either as a parametric or nonparametric function. We refer the reader to the books of Ferraty and Vieu (2006) and Ferraty and Romain (2010) for an overview of the nonparametric approach, with more advanced results that can be found in Ferraty and Vieu (2002) and Ferraty et al. (2007). In the parametric way, the most extensively studied model is the functional linear model

$$y = \langle x, \beta \rangle + \varepsilon, \quad (2)$$

* Corresponding author. Tel.: +86 431 85099978.

E-mail addresses: wanggc023@nenu.edu.cn (G. Wang), nlin@math.wustl.edu (N. Lin), bxzhang@nenu.edu.cn (B. Zhang).

where $r(\cdot) = \langle \beta, \cdot \rangle$ for a certain functional regression coefficient $\beta \in L^2([a, b])$, x is a zero-mean random variable such that $E(\|x\|^2) < \infty$, and ε satisfies $E(\varepsilon) = 0$, $E(\varepsilon^2) < \infty$ and $E(x\varepsilon) = 0$. There are two major methods to estimating β . One is the regularization method (see Cardot et al., 2003; Ramsay and Silverman, 2002, 2005; Yuan and Cai, 2010), and the other is based on functional principal component analysis (see Cai and Hall, 2006; Ferraty et al., 2011; Manteiga and Calvo, 2011).

When interpretability of the model is a more important study, the parametric approach, especially functional linear regression model (2), is generally preferable. However, the linearity assumption may not be satisfied and some transformation on the response variable y becomes necessary. For example, consider the presence of heteroscedasticity where the error term $\varepsilon = \sigma(x)\tilde{\varepsilon}$ with $\tilde{\varepsilon}$ independent of x and $E(\tilde{\varepsilon}) = 0$ and $\text{var}(\tilde{\varepsilon}) = 1$, and $\sigma(x)$ is a real function satisfying $E(\sigma(x)^2) < \infty$. Using model (1) and (2) can lead to invalid results (see Crambes et al., 2008; Delaigle et al., 2009). Currently, when transformation on the response y is necessary, the transformation function needs to be prespecified with a known form. For example, in functional nonparametric regression, transformation models have been studied by Crambes et al. (2008) and Ferraty et al. (2010) with the transformation function being prespecified with a known form. Crambes et al. (2008) focused on robust nonparametric estimation of r in model (1), and Ferraty et al. (2010) proved the uniform almost complete convergence rate of the nonparametric estimate. In functional linear regression analysis, Ramsay and Dalzell (1991) used logarithm transformation when analyzing the Canadian climate data. However, for complex data, a proper transformation is generally difficult to pre-conceive and the users will have to rely on a cumbersome trial-and-error process, and such efforts are easily futile if the transformation is not of a simple parametric form.

In this article, our goal is to develop a systematic and flexible approach to decide such transformations in the context of functional linear regression by simultaneously estimating the transformation function and the functional regression coefficients. Motivated by He and Shen (1997), we propose a nonparametric approach called the functional linear regression after spline transformation (FLIRST). That is, we extend (2) as

$$h(y) = \langle x, \beta \rangle + \varepsilon, \quad (3)$$

where $h(\cdot)$ is an unknown smooth function approximated by a quadratic B -spline, and estimated together with β . Since the functional coefficient β in model (3) is identifiable only up to a scale multiplication, the real interest will be in estimating its normalized version. We propose an innovative algorithm called mixed-data canonical correlation analysis (MDCCA) to consistently estimate the functional coefficient and the transformation function. MDCCA is analogous to the canonical correlation analysis between two multivariate samples, but is between a multivariate sample and a set of functional data. Knots selection for the B -spline is then solved as a model selection problem.

Our approach is closely related to the regularized functional sliced inverse regression (RFSIR) of Ferré and Villa (2006). They considered the following model

$$y = g(\langle x, \beta_1 \rangle, \dots, \langle x, \beta_K \rangle, \varepsilon), \quad (4)$$

where g is an unknown link function, and β_j 's are linearly independent functions in H^2 , the subspace of $L^2([a, b])$ that contains the functions with a squared integrable second derivative. It is worth noting that, though FLIRST and RFSIR are related, their goals are different. FLIRST focuses more on finding the proper transformation on the response variable, while RFSIR focuses more on dimension reduction and does not provide an estimate of the link function g . Our FLIRST model (3) can be viewed as a special case of (4) when $K = 1$ plus some mild constraints on $g(\cdot)$. We will show that, given a linear condition, FLIRST and RFSIR are equivalent under model (4) with $K = 1$. When the linear condition does not hold, FLIRST and RFSIR no longer provide consistent estimates of the functional parameter, but the FLIRST estimate still gives the projection of the functional predictor that best correlates with the transformed response.

The rest of this paper is organized as follows. In Section 2, we first describe the MDCCA algorithm and then propose the FLIRST methodology as an application of MDCCA to functional linear regression transformation model. Theoretical properties of the FLIRST method are also given. In Section 3, we perform simulation studies to compare FLIRST with other methods. Real data examples are then presented in Section 4, and we conclude the paper and provide some discussions in Section 5.

2. Functional linear regression after spline transformation

In finding transformations for the response in linear regression, He and Shen (1997) showed that canonical correlation analysis (CCA) can be used to obtain consistent estimates of the regression coefficients when the transformation function is approximated by a spline. We extend this idea to functional linear regression and will need a 'CCA' between a multivariate sample and a set of functional predictors, which we call mixed-data CCA (MDCCA).

2.1. Mixed-data canonical correlation analysis

Let \mathbf{y} denote a p -dimensional random vector and x be a random curve in $L^2([a, b])$. To keep notation simple, we define the covariance function of x by $\Gamma(s, t) = E(x(s)x(t))$ and then we define Γg as the function $\Gamma g(s) = \langle \Gamma(s, t), g \rangle = \int_a^b \Gamma(s, t)g(t)dt$ for $g \in L^2([a, b])$. For the moment, we shall focus on the leading *canonical variate*. That is, we want to find a vector \mathbf{b} and a function β that maximize the correlation coefficient between $\mathbf{b}^T \mathbf{y}$ and $\langle x, \beta \rangle$. Let $\gamma(\mathbf{b}, \beta)$ to be the squared correlation coefficient of $\mathbf{b}^T \mathbf{y}$ and $\langle x, \beta \rangle$, i.e.

متن کامل مقاله

دریافت فوری ←

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