Journal of Multivariate Analysis 161 (2017) 68–82

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A class of functional partially linear single-index models

Hui Ding [a,b], Yanghui Liu [a], Wenchao Xu [a], Riquan Zhang [a,*]

[a] School of Statistics, East China Normal University, Shanghai, 200241, PR China
[b] School of Mathematics and Finance, Chuzhou University, Chuzhou, 239000, PR China

ARTICLE INFO

Article history:
Received 6 June 2016
Available online 22 July 2017

Keywords:
Convergence rate
Functional data analysis
Partial linear
Prediction
Single-index

ABSTRACT

The functional linear regression model is a useful extension of the classical linear model. However, it assumes a linear relationship between the response and functional covariates which may be invalid. For this reason, we generalize this model to a class of functional partially linear single-index models. In this paper, we propose a profile least squares approach combined with local constant smoothing for estimating the slope function and the link function in the new model. We demonstrate that our methods enable prediction of the link function and estimation of the slope function with polynomial convergence rates. The convergence rate of prediction of the whole model is also established. Monte Carlo simulation studies show an excellent finite-sample performance. A real data example about average yield of oats in Saskatchewan, Canada is used to illustrate our methodology.

1. Introduction

Functional data analysis has become more and more popular owing to its adaptability to problems which are hard to deal with in the framework of scalar and vector observations. Samples of functional data consist of random functions, each of which is viewed as one sampling unit. Functional data are inherently infinite-dimensional and are rich in information. Hence, functional data analysis has found applications in many subject areas, such as biology, medical sciences, meteorology, econometrics, finance, chemometrics and geophysics.

One of the most important tools in functional data analysis is functional regression. Functional linear regression, which is the simplest form of functional regression, has been frequently researched and used in practice. Formally, a functional linear regression model can be written as

$$Y = \int_I \alpha(t)X(t)\mathrm{d}t + \varepsilon,$$

where $X(t)$ is a functional covariate defined on a compact interval I, Y is a scalar response variable, $\alpha(t)$ is an unknown slope function and ε is an error term. There is a great amount of the literature on functional linear regression. For example, Ramsay and Silverman [11] introduced the above functional linear model to express the relationship between a functional predictor measured across a dense grid of regularly spaced time points and a scalar response variable. Cardot et al. [2,3] and Hall and Horowitz [7] proposed various estimation methods for the functional linear model using functional principal component analysis or smoothing techniques and investigated the asymptotic behavior of the proposed estimators. Yao et al. [20] proposed an estimation method for the functional linear model when the functional predictor is measured with error at irregularly spaced time points.

* Corresponding author.
E-mail address: zhangriquan@163.com (R. Zhang).

http://dx.doi.org/10.1016/j.jmva.2017.07.004
0047-259X/

While functional linear regression has been widely used in many fields, it assumes a linear relationship between a scalar response and functional covariates. In practice, the linear assumption may not be effective. Thus, Chen et al. [5] proposed a single-index functional model

$$Y = g\left\{\int_I \alpha(t)X(t)\mathrm{d}t\right\} + \varepsilon,$$

where the nonparametric link function g is unknown. They used the method of least squares together with a local constant or local linear smoothing technique to estimate the link function g and the slope function $\alpha(t)$. They also proved that their methods enable prediction with polynomial convergence rates and applied the model to predict the fat content of meat specimens from a spectrometric curve.

However, the single-index functional model assumes a single functional covariate and may not be sufficient to explain the variation of the scalar response variable Y via the functional covariate $X(t)$. In practice, it is commonly the case that a scalar response is related to two functional covariates $X(t)$ and $Z(t)$. For instance, as we discuss in Section 5, average yield of oats depends not only on temperature but also on precipitation. Moreover, the relationships between the response and functional covariates may not be all linear, but we want to retain the ease of interpretation of slope functions $\alpha(t)$ and $\beta(t)$ for two functional covariates $X(t)$ and $Z(t)$.

Here, we borrow an idea from Carroll et al. [4] by making the linear combination $\int_{I_1} \alpha(t)X(t)\mathrm{d}t$ enter the model via a nonparametric link function, and $\int_{I_2} \beta(t)Z(t)\mathrm{d}t$ enter the model as an offset. This leads to a class of functional partially linear single-index models, viz.

$$Y = g\left\{\int_{I_1} \alpha(t)X(t)\mathrm{d}t\right\} + \int_{I_2} \beta(t)Z(t)\mathrm{d}t + \varepsilon.$$

To the best of our knowledge, the above functional partially linear single-index model has not been studied in the scientific literature. The model is flexible in practice, which motivates us to investigate it. In our article, the slope functions $\alpha(t)$ and $\beta(t)$ are still potentially of interest, but the primary focus of this paper is prediction rather than estimation of slope functions. Our theory addresses directly the prediction problem.

Prediction based on functional covariates arose in many applications over the last ten years. Thus, the functional regression model has attracted considerable attention on prediction. Cai and Hall [1] investigated prediction of response in the setting of the functional linear model. They used a principal component analysis approach to obtain an estimator of the slope function and proved that the method can enable prediction to achieve the minimax convergence rate under some conditions. Crambes et al. [6] considered prediction of response in the functional linear model. They used a smoothing spline estimator for the functional slope parameter based on a slight modification of the usual penalty. They also showed that the convergence rates of the prediction are optimal over a large class of possible slope functions and distributions of the predictive curves. Chen et al. [5] studied prediction of response in the framework of the single-index functional regression model. They proposed a new method to estimate parameters and the link function. They also proved that their methods enable prediction with polynomial convergence rates. Shin and Lee [13] considered prediction of response in a partially functional linear model. In order to obtain estimators of parameters, they investigated two different approaches. One is with functional principal component regression and the other is related to functional ridge regression based on the Tikhonov regularization. They also demonstrated that two different methods can achieve the same convergence rate of the mean squared prediction error under appropriate assumptions.

In this paper, we use a profile least squares approach combined with a local constant smoothing technique to estimate the slope functions $\alpha(t)$, $\beta(t)$ and the link function g of the functional partially linear single-index model. Our aim lies in the prediction of the scalar response. Specifically, for the criterion of mean squared prediction error, we adopt a similar form as Chen et al. [5], i.e., mean squared prediction error is given by

$$n^{-1}\sum_{i=1}^{n}\left[\hat{g}\left\{\int_{I_1} \hat{\alpha}(t)X_i(t)\mathrm{d}t\right\} + \int_{I_2} \hat{\beta}(t)Z_i(t)\mathrm{d}t - \left[g\left\{\int_{I_1} \alpha(t)X_i(t)\mathrm{d}t\right\} + \int_{I_2} \beta(t)Z_i(t)\mathrm{d}t\right]\right]^2.$$

Under some regularity conditions, we show that the estimator of the link function g converges to g at a polynomial rate and also demonstrate that our methods enable prediction with polynomial convergence rates.

The article is organized as follows. Section 2 describes our model and our estimation method. Sections 3 and 4 present asymptotic properties and finite-sample performance of the proposed estimators respectively. Section 5 provides an application to average yield of oats data from Saskatchewan, Canada, and a summary of our findings is given in Section 6. Technical arguments are relegated to an Appendix.

2. Model and estimation

2.1. The model

We first consider the following functional partially linear single-index model

$$Y = g_1\left\{\int_{I_1} \alpha(t)X(t)\mathrm{d}t\right\} + \int_{I_2} \beta(t)\tilde{Z}(t)\mathrm{d}t + \varepsilon, \tag{1}$$