Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Varying-coefficient partially functional linear quantile regression models

Ping Yu [a,b], Jiang Du [a,c], Zhongzhan Zhang [a,c,*]

[a] College of Applied Sciences, Beijing University of Technology, Beijing, 100124, China
[b] School of Mathematics and Computer Science, Shanxi Normal University, Linfen, 041000, China
[c] Collaborative Innovation Center on Capital Social Construction and Social Management, 100124, China

A R T I C L E I N F O

Article history:
Received 6 August 2016
Accepted 16 February 2017
Available online xxxx

AMS 2000 subject classifications:
primary 62G08
secondary 62G20

Keywords:
Functional data analysis
Varying-coefficient quantile model
Functional linear quantile regression model
B-splines
Functional principal component analysis

A B S T R A C T

In this paper, we introduce a new varying-coefficient partially functional linear quantile regression model, which combines varying-coefficient quantile regression model with functional linear quantile regression model. The functional principal component basis and regression splines are employed to estimate the slope function and varying-coefficient functions, respectively, and the convergence rates of the estimators are obtained under some regularity conditions. Simulations and an illustrative real example are presented.

1. Introduction

Functional data analysis (FDA) is a topic of growing interest in statistics and is applied in econometrics, chemometrics, biology, and other fields, see, for example, Ferraty and Vieu (2006) and Ramsay and Silverman (2002, 2005). In the simplest setting, the functional predictor and the scalar response are related by a linear operator. For a scalar response Y and a smooth square integrable random predictor process $X(\cdot)$ on a compact support $\mathscr{T}$, the linear relationship between Y and X is expressed as

$$E(Y|X) = \alpha + \int_{\mathscr{T}} \beta(t)X(t)dt, \tag{1}$$

where α is the intercept, slope function $\beta(\cdot)$ is assumed to be square integrable on $\mathscr{T}$. We note that there are numerous works on the functional linear regression model, see, e.g., Cai and Hall (2006), Cai and Yuan (2012), Cardot, Ferraty, and Sarda (1999), Crambes, Kneip, and Sarda (2009), Hall and Horowitz (2007), Lei (2014), Yao, Müller, and Wang (2005) and references therein.

Since the functional linear model is too restrictive on the regression relation, several extensions of functional linear models have been proposed. These include semi-functional partial linear regression (Aneiros-Pérez & Vieu, 2006; Kong,

* Corresponding author at: College of Applied Sciences, Beijing University of Technology, Beijing, 100124, China.
E-mail address: zzhang@bjut.edu.cn (Z. Zhang).

http://dx.doi.org/10.1016/j.jkss.2017.02.001
1226-3192/

Xue, Yao, & Zhang, 2016; Lian, 2011; Shin, 2009), varying-coefficient functional models (Cardot & Sarda, 2008; Wu, Fan, & Müller, 2010), generalized functional linear models (Müller & Stadtmüller, 2005), functional additive models (Müller & Yao, 2008), and so on. In addition, varying coefficient models (Fan & Zhang, 2008; Hastie & Tibshirani, 1993) are useful extensions of the classical linear models, which extend the applications of local regression techniques from one dimensional to multidimensional setting. Thus, Peng, Zhou, and Tang (2015) introduced varying-coefficient partially functional linear regression models combining the functional linear model with the popular varying coefficient model for scalar variables.

However, the above-mentioned references are focused on mean regression, and relatively few studies are from quantile regression perspective. Compared with classical conditional mean regression, quantile regression has at least three advantages. First, quantile regression, in particular median regression, provides an alternative and complement to mean regression while being resistant to outliers in responses; in addition, quantile regression is more efficient than mean regression when the error follows a distribution with heavy tails. Second, quantile regression can give a more complete picture on how the responses are affected by covariates. It is particularly useful when upper or lower or all quantiles are of interest. This is attractive in economic and actuaries, where the tail behavior of the response conditional on covariates is often concerned with. Third, quantile regression is capable of dealing with heteroscedasticity, the situation in which variances depend on certain covariates. For a broad review on quantile regression, the reader may refer to Koenker (2005) and Kong, Maity, Hsu, and Tzeng (2016). We also note that there are a few of references about FDA based on quantile regression, see, e.g., Cardot, Crambes, and Sarda (2005), Kato (2012) and Lu, Du, and Sun (2014). In this paper, we consider varying-coefficient partially functional linear quantile regression model, which includes not only the model in Kim (2007) but also the models in Kato (2012) and Lu et al. (2014) as special cases.

Let Y be a real-valued random variable defined on a probability space $(\Omega, \mathscr{B}, P)$, U and $\boldsymbol{Z} = (Z_1, \dots, Z_p)^T$ be independent one-dimensional and p-dimensional vectors of explanatory variables defined on the same probability space, respectively. We suppose that U ranges over a non-degenerate compact interval, say $[0, 1]$. Also, let $\{X(t) : t \in \mathscr{T}\}$ be a zero mean, second-order stochastic process valued in $H = L^2(\mathscr{T})$, the set of all square integrable functions on $\mathscr{T}$ with inner product $\langle x, y\rangle = \int_{\mathscr{T}} x(t)y(t)dt, \forall x, y \in H$ and norm $\|x\| = \langle x, x\rangle^{1/2}$. At a given quantile level $\tau \in (0, 1)$, the linear relationship between Y and $\{X, U, \boldsymbol{Z}\}$ is assumed as

$$Y = \int_{\mathscr{T}} \alpha_\tau(t)X(t)dt + \boldsymbol{Z}^T\boldsymbol{\beta}_\tau(U) + \epsilon_\tau, \tag{2}$$

where $\alpha_\tau(t) \in H$ and $\boldsymbol{\beta}_\tau(U) = (\beta_{1\tau}(U), \dots, \beta_{p\tau}(U))^T$ are unknown smooth functions of U, ϵ_τ is a random error whose τth quantile conditional on $(X, \boldsymbol{Z}, U)$ equals zero. Without loss of generality, we suppose throughout that $\mathscr{T} = [0, 1]$. For the ease of presentation, we will suppress τ in $\alpha_\tau(t)$, $\boldsymbol{\beta}_\tau(U)$ and ϵ_τ in model (2) wherever clear from the context. Clearly, model (2) generalizes both the varying-coefficient quantile regression (Kim, 2007) and the functional linear quantile regression (Kato, 2012) which correspond to the cases $\alpha(t) = 0$ and $\boldsymbol{\beta}(U) = \boldsymbol{0}$, respectively. Note that model (2) becomes the partial functional linear quantile regression (Lu et al., 2014) when $\boldsymbol{\beta}(U) = \boldsymbol{\theta}$.

The rest of paper is organized as follows. In Section 2, we introduce the functional principal component analysis (FPCA) and spline-based method to estimate the slope function and varying-coefficient functions, respectively. Section 3 gives the rates of convergence of the estimators. Some simulation results and an example are presented in Sections 4 and 5, respectively. We conclude this article with a brief discussion in Section 6. All proofs are given in Appendix.

2. Estimation methods

In this section, we describe how to estimate the slope function and varying-coefficient functions. As is discussed in Delaigle and Hall (2012), the FPCA is a benchmark basis in functional data. Furthermore, regression splines have some desirable properties in approximating a smooth function, and often provide good approximations with small number of knots. For these reasons, we approximate the slope function by FPCA basis and varying-coefficient functions by regression splines in the following.

Firstly, let $(X_i, \boldsymbol{Z}_i, Y_i, U_i)$, $i = 1, \dots, n$ be i.i.d. realizations of $(X, Y, \boldsymbol{Z}, U)$, which are generated from model (2), i.e.

$$Y_i = \int_0^1 \alpha(t)X_i(t)dt + \beta_1(U_i)Z_{i1} + \cdots + \beta_p(U_i)Z_{ip} + \epsilon_i. \tag{3}$$

Define the covariance function of X and its empirical version respectively as

$$C(t, s) = \mathrm{Cov}(X(t), X(s)), \qquad \hat{C}(t, s) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)X_i(s).$$

The covariance function C defines a linear operator which maps a function f to Cf given by $Cf(s) = \int C(t, s)f(t)dt$. We shall assume that the linear operator with kernel C is positive definite. By the Mercer's Theorem, we have

$$C(t, s) = \sum_{i=1}^{\infty} \lambda_i v_i(t)v_i(s), \qquad \hat{C}(t, s) = \sum_{i=1}^{\infty} \hat{\lambda}_i\hat{v}_i(t)\hat{v}_i(s),$$