STAPRO: 8286 Model 3G pp. 1–10 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Semi-functional partial linear quantile regression☆

Hui Ding[a,b], Zhiping Lu[a], Jian Zhang[c], Riquan Zhang[a,*]

[a] School of Statistics, East China Normal University, Shanghai, 200241, PR China
[b] School of Economics, Nanjing University of Finance and Economics, Nanjing, 210023, PR China
[c] School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7NF, UK

ARTICLE INFO

Article history:
Received 31 August 2017
Received in revised form 3 July 2018
Accepted 4 July 2018
Available online xxxx

MSC:
primary 62G05
secondary 62G20

Keywords:
Functional data analysis
Partial linear
Quantile regression

ABSTRACT

Semi-functional partial linear model is a flexible model in which a scalar response is related to both functional covariate and scalar covariates. We propose a quantile estimation of this model as an alternative to the least square approach. We also extend the proposed method to kNN quantile method. Under some regular conditions, we establish the asymptotic normality of quantile estimators of regression coefficient. We also derive the rates of convergence of nonparametric function. Finite-sample performance of our estimation is compared with least square approach via a Monte Carlo simulation study. The simulation results indicate that our method is much more robust than the least square method. A real data example about spectrometric data is used to illustrate that our model and approach are promising.

1. Introduction

Over the last two decades, technological progress in many subject areas have produced a large number of continuous data with curves or images as the units of observation. Functional data analysis (FDA) encompasses the statistical methodology for such data and has been prevailed. See Müller (2005), Cuevas (2014), Morris (2015), Wang et al. (2016) and Goia and Vieu (2016) for systematic reviews on this subject. The recent monographs by Horváth and Kokoszka (2012) and Hsing and Eubank (2015) offer some mathematical theories of functional data. As the important tool of FDA, functional regression aim to model the relationship between functional (scalar) response and functional (scalar) covariates. Researchers are increasingly interested in functional regression models. See Greven and Scheipl (2017) for a short survey on this field. It is noteworthy that semiparametric functional regression models offer a well-balanced mixture of parametric models and nonparametric models. Semiparametric functional regression models keep flexibility of parametric regression models and overcome sensitivity to dimensional effects of nonparametric approaches. See Goia and Vieu (2014) for a short survey. Semi-functional partial linear regression model is an important semiparametric functional regression model. It can be expressed as $Y = m(X) + \boldsymbol{Z}^{\top}\boldsymbol{\beta} + \varepsilon$, where X is functional covariate that taking values in semi-metric space $\mathscr{F}$, $m : \mathscr{F} \to \mathscr{R}$ is a unknown smooth function, $\boldsymbol{Z}$ is p-dimensional random vector of scalar covariate, $\boldsymbol{\beta}$ is unknown coefficient of scalar covariate, ε is a random error. The model has been widely used in many fields. Aneiros-Pérez and Vieu (2006) proposed profile least square estimation method and derived the asymptotic performances of proposed estimators. Aneiros-Pérez and Vieu (2008) extended the model to time series area. Ling et al. (2017) proposed a k-nearest-neighbours (kNN) procedure and derived the

☆ The research was supported in part by National Natural Science Foundation of China (11571112, 11701360), Program of Shanghai Subject Chief Scientist (14XD1401600), the 111 Project of China (B14019), Research Innovation Program for ECNU Graduates (ykc17083), the Natural Science Foundation of the Anhui Provincial Department of Education (KJ2017A424, KJ2018A0424) and the Anhui Province Natural Science Foundation (1508085QA14).
* Corresponding author.
E-mail address: zhangriquan@163.com (R. Zhang).

https://doi.org/10.1016/j.spl.2018.07.007

Please cite this article in press as: Ding H., et al., Semi-functional partial linear quantile regression. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.07.007.

asymptotic performances of kNN estimators. Aneiros et al. (2015) extend the model to high-dimensional framework. They proposed penalized least-squares method to study the problem of variable selection and derived an oracle property.

The above-mentioned references are all focused on mean regression. It is known that mean regression is sensitive to outliers. Quantile regression is usually recognized as an alternative to mean regression. Quantile regression is more robust than mean regression. There is few literatures on quantile-regression-based estimation procedures in the functional regression model. Cardot et al. (2005) proposed a spline estimator for functional linear quantile regression model. Chen and Müller (2012) studied a estimation method for conditional quantile analysis in the generalized functional regression framework. Kato et al. (2012) studied quantile estimation in functional linear quantile regression model. In this paper, we study quantile regression of semi-functional partial linear model. To the best of our knowledge, this method has not been researched in the scientific literature. Since the model is flexible in practice, quantile regression method is urgently needed, which motivates us to investigate quantile regression of estimation of the model.

In this paper, we use quantile regression method to estimate the nonparametric function and regression coefficient of the model. We also extend the proposed method to kNN quantile method. Under some regular conditions, we establish the asymptotic normality of estimators of regression coefficient and derive the rates of convergence of nonparametric function. A Monte Carlo simulation and an application to spectrometric data show the advantages of our proposed method.

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 spectrometric data. Concluding remarks are provided in Section 6. Technical proofs are given in an Appendix.

2. Model and estimation

2.1. Model

Given quantile level $\tau \in (0, 1)$, we consider the following semi-functional partial linear quantile regression model

$$Y = m_\tau(X) + \boldsymbol{Z}^\top \boldsymbol{\beta}_\tau + \varepsilon_\tau, \tag{1}$$

where X is functional covariate that taking values in semi-metric space $\mathscr{F}$, and we denote the associated semi-metric by $d(\cdot,\cdot)$, $m_\tau : \mathscr{F} \to \mathscr{R}$ is a unknown smooth function, $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^\top$ are p-dimensional random vector of scalar covariates, $\boldsymbol{\beta}_\tau = (\beta_{1\tau}, \ldots, \beta_{p\tau})^\top$ are unknown coefficients of scalar covariates, ε_τ is a random error whose τth quantile conditional on $(\boldsymbol{Z}, X)$ being zero.

2.2. Estimation

Suppose that $\{(Y_i, X_i, \boldsymbol{Z}_i), i = 1, \ldots, n\}$ is a random sample generated from model (1). We estimate coefficients $\boldsymbol{\beta}_\tau$ and function $m_\tau(\cdot)$ in model (1), by minimizing the following quantile loss function

$$\sum_{i=1}^{n} \rho_\tau \left(Y_i - m_\tau(X_i) - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_\tau \right), \tag{2}$$

where $\rho_\tau(s) = s\{\tau - I(s < 0)\}$.

Obviously, (2) contains both nonparametric and parametric component. And they can be estimated by different rates of convergence, so we use three-stage procedure. In the first stage, we apply the local constant weighted quantile smoothing technique to get an initial estimators. That is, we obtain an initial estimators of $m_\tau(X_j)$ and $\boldsymbol{\beta}_\tau$ by minimizing the following local weighted quantile loss function

$$\sum_{i:i\neq j} \rho_\tau \left(Y_i - a_{\tau j} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_\tau \right) K_{h_0}(d(X_i, X_j)), \tag{3}$$

where $K_{h_0}(\cdot) = K(\cdot/h_0)/h_0$ and $K(\cdot)$ is a kernel function and h_0 is a bandwidth. For convenience, we denote the initial estimators of $a_{\tau j}, \boldsymbol{\beta}_\tau$ by $\tilde{a}_{\tau j}, \tilde{\boldsymbol{\beta}}_\tau$. In the second stage, we further improve the efficiency of $\tilde{\boldsymbol{\beta}}_\tau$. Specifically, we derive the final estimator of $\boldsymbol{\beta}_\tau$ by minimizing the following quantile loss function

$$\sum_{j=1}^{n} \rho_\tau \left(Y_j - \tilde{a}_{\tau j} - \boldsymbol{Z}_j^\top \boldsymbol{\beta}_\tau \right). \tag{4}$$

Denote the final estimator of $\boldsymbol{\beta}_\tau$ by $\hat{\boldsymbol{\beta}}_\tau$. In the third stage, we obtain the final estimator of $m_\tau(x)$. More concretely, we obtain the final estimator of $m_\tau(x)$ by minimizing the following local weighted quantile loss function

$$\sum_{i=1}^{n} \rho_\tau \left(Y_i - a_\tau - \boldsymbol{Z}_i^\top \hat{\boldsymbol{\beta}}_\tau \right) K_h(d(X_i, x)). \tag{5}$$

Denote the final estimators of a_τ by $\hat{a}_\tau$. Evidently, $\hat{a}_\tau$ are the final estimator of $m_\tau(x)$.

Please cite this article in press as: Ding H., et al., Semi-functional partial linear quantile regression. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.07.007.