Computational Statistics and Data Analysis 62 (2013) 181–187

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

A partial spline approach for semiparametric estimation of varying-coefficient partially linear models

Young-Ju Kim

Department of Statistics, Kangwon National University, 192-1 Hyoja-dong, Chuncheon 200-701, Republic of Korea

ARTICLE INFO

Article history:
Received 2 March 2012
Received in revised form 28 December 2012
Accepted 7 January 2013
Available online 11 January 2013

Keywords:
Bayesian confidence interval
Partial spline
Partially linear
Penalized likelihood
Smoothing spline
Varying coefficients

ABSTRACT

A semiparametric method based on smoothing spline is proposed for the estimation of varying-coefficient partially linear models. A simple and efficient method is proposed, based on a partial spline technique with a lower-dimensional approximation to simultaneously estimate the varying-coefficient function and regression parameters. For interval inference, Bayesian confidence intervals were obtained based on the Bayes models for varying-coefficient functions. The performance of the proposed method is examined both through simulations and by applying it to Boston housing data.

1. Introduction

Consider observations $y_i = \eta(\mathbf{x}_i) + \epsilon_i$, $i = 1, \ldots, n$, where y_i is the response variable, $\mathbf{x}_i$ are predictors, and ϵ_i are i.i.d. random errors with $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$. Classical parametric regression would assume η to be of the form $\eta(\mathbf{x}, \beta)$, where β parameters are estimated from the data. The parametric assumption can be relaxed by allowing η to vary in a high (possibly infinite)-dimensional function space and this leads to nonparametric function estimation approach. A major problem when applying nonparametric methods is that they suffer the "curse of dimensionality". Many attempts have been made to overcome this using various structural modeling techniques, including the generalized additive model and varying-coefficient models. The application of the semiparametric approach to the varying-coefficient model leads to the varying-coefficient partially linear model (VCPLM) in which some of the predictors have flexible coefficient functions whereas others have constant coefficients. For predictors $\mathbf{x}$ and $\mathbf{z}$, the VCPLM can be expressed as

$$y_i = \mathbf{x}_i^T \eta(\mathbf{u}_i) + \mathbf{z}_i^T \beta + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{x}_i$ is of dimension $p_1 \times 1$, $\mathbf{z}_i$ is of dimension $p_2 \times 1$, $\mathbf{u}_i = (u_{i1}, \ldots, u_{ip_1})$ is a vector of predictors, $\eta(\cdot) = (\eta_1(\cdot), \ldots, \eta_{p_1}(\cdot))^T$ is a $p_1 \times 1$ vector of unknown smooth functions, and $\boldsymbol{\beta}$ is a $p_2 \times 1$ vector of unknown parameters. The varying-coefficient components in (1) provide a way to determine how the different values of covariate $\mathbf{u}$ influence the effect of $\mathbf{x}$ on $\mathbf{y}$ by allowing the coefficient of $\mathbf{x}$ to be a function of $\mathbf{u}$, which makes the model flexible.

The VCPLM (1) covers many common regression models. When $\beta = 0$ and $\mathbf{x} = 1$, the estimation of the model in (1) reduces to classical nonparametric function estimation problem. When $p_1 = 1$ and $\mathbf{x} = 1$, the model reduces to the partially linear model (Wahba, 1984; Gu, 2002; Kim, 2010), while this becomes the varying-coefficient model when $\beta = 0$ (Hastie and Tibshirani, 1993; Fan and Zhang, 1999; Fan et al., 2003; Eubank et al., 2004). The varying-coefficient model have

E-mail address: ykim7stat@kangwon.ac.kr.

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2013.01.006

been applied in many different applications, including generalized linear models, time series analysis, and longitudinal data analysis (e.g. Cai et al., 2000; Chiang et al., 2001; Huang et al., 2004; Senturk and Muller, 2008, and references therein).

The model (1) has been studied over decades by using various nonparametric estimation methods; kernel-based methods (Zhang et al., 2002; Xia et al., 2004; Fan and Huang, 2005; Li and Racine, 2010) and spline-based methods (Hoover et al., 1998; Lu et al., 2008; Wang and Ke, 2009; Ahmad et al., 2010). The spline-based methods are found to be more attractive for its flexibility to involve multiple smoothing parameters, while they often encounter computational challenge in practice since the number of spline basis functions can be large (Hoover et al., 1998; Krafty et al., 2008).

In this paper we propose a simple and efficient method based on partial spline techniques with a lower-dimensional approximation to estimate both the varying-coefficient function and regression parameters. The smoothing parameter was selected using modified generalized cross-validation (GCV). For interval inference, Bayesian confidence intervals were obtained based on the Bayes models for varying-coefficient functions.

The paper is organized as follows. Section 2 describes the partial spline in a lower-dimensional approximating function space and the Bayes model for the estimator. Section 3 presents the computation method and its algorithm employing smoothing parameter selection methods. Section 4 reports the numerical results from simulation examples, and Section 5 describes the application of the proposed method to Boston housing data. Some concluding remarks are given in Section 6.

2. The model

2.1. Partial splines

A partially linear model is described by

$$y_i = \eta(u_i) + \mathbf{z}_i^T \beta + \epsilon_i, \quad i = 1, \ldots, n. \tag{2}$$

Partial splines estimate η and β in (2) are obtained by minimizing the following penalized least squares functional:

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - \eta(u_i) - \mathbf{z}_i^T\beta\}^2 + \lambda J(\eta), \tag{3}$$

where $J(\eta)$ is a roughness penalty functional, and smoothing parameter λ controls the trade-off between the lack of fit and the roughness of η. The proper selection of smoothing parameter determines the performance of the estimator. The minimizer of (3) is in infinite-dimensional space $\mathcal{H} \subseteq \{f : J(f) < \infty\}$. Specifically, the minimizer of (3) lies in a Hilbert space $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_J$, where $\mathcal{H}_0 = \mathcal{H}_{00} \oplus \mathcal{N}_J$, $\mathcal{H}_{00} = \text{span}\{\psi_l, l = 1, \ldots, p_2\}$, $\{\psi_l\}$ are rows of the matrix $\mathbf{z}$, $\mathcal{N}_J = \{\eta : J(\eta) = 0\}$ is the null space of $J(\eta)$, and space $\mathcal{H}_J$ is a reproducing kernel Hilbert space(RKHS) with $J(\eta)$ as the square norm. Note that a space $\mathcal{H}$ in which evaluation functional $[x]f = f(x)$ is continuous is called an RKHS, which possesses a reproducing kernel (RK) $R(\cdot,\cdot)$, a nonnegative definite function satisfying $\langle R(x,\cdot), f(\cdot)\rangle = f(x), \forall f \in \mathcal{H}$, where $\langle\cdot,\cdot\rangle$ is the inner product in $\mathcal{H}$. Letting $J(f) = \int_0^1 \ddot{f}^2 dt$ on $\mathcal{T} = [0, 1]$ yields the popular cubic splines with $\mathcal{N}_J = \text{span}\{1, k_1(t)\}$, where $k_1(t) = t - 0.5$. In $\mathcal{H}_J = \{f : \int_0^1 f dt = \int_0^1 \dot{f} dt = 0, J(f) < \infty\}$ with $J(f)$ as the square norm, one has the RK $R_J(t_1, t_2) = k_2(t_1)k_2(t_2) - k_4(t_1 - t_2)$, where $k_\nu = B_\nu/\nu!$ are scaled Bernoulli polynomials. Wahba (1990) and Gu (2002) provide details of the RKHS and its properties.

A data-adaptive lower-dimensional approximation can be used in penalized likelihood methods, as originally proposed by Gu and Kim (2002) for regression. They showed that the convergence rate of the minimizer of the penalized likelihood functional in $\check{\mathcal{H}}_q = \mathcal{N}_J \oplus \mathcal{H}_J$ was the same as that in the lower-dimensional function space $\check{\mathcal{H}}_q = \mathcal{N}_J \oplus \text{span}\{R_J(w_j, \cdot), j = 1, \ldots, q\}$, where $\{w_j\}$ are random subsets of $\{u_i, i = 1, \ldots, n\}$, as long as $q \asymp n^{2/(pr+1)+\epsilon}$, where for some $p \in [1, 2]$, $r > 1$, $\epsilon > 0$ is arbitrary. Here p represents the smoothness of the true function and $p = 2$ is used under the assumption that the true function is sufficiently smooth. We extend their results to the model (1) so as to speed up the computation of the function estimators without any loss of performance. The constant r characterizes the smoothness of the model and $r = 4$ is used for the cubic spline.

Letting $h(\mathbf{z}, u) = \eta(u) + \mathbf{z}^T\beta$ in (2) and using an argument similar to that of Wahba (1990) for partial splines, the minimizer of (3) in $\mathcal{H}_q = \mathcal{H}_0 \oplus \text{span}\{R_J(w_j, \cdot), j = 1, \ldots, q\}$ can be written as

$$h = \sum_{\nu=1}^{m} d_\nu \phi_\nu(u) + \sum_{i=1}^{q} c_i R_J(w_i, u) + \sum_{l=1}^{p_2} \beta_l \psi_l, \tag{4}$$

where $\{\phi_\nu\}$ is a basis of null space $\mathcal{N}_J$. Then, the problem becomes to find β, $\mathbf{c}$, and $\mathbf{d}$ so as to minimize

$$(\mathbf{y} - S\mathbf{d} - R\mathbf{c} - \Sigma\beta)^T(\mathbf{y} - S\mathbf{d} - R\mathbf{c} - \Sigma\beta) + \lambda\mathbf{c}^T Q\mathbf{c},$$

where S is $n \times m$ with (i, ν)th entry $\phi_\nu(u_i)$, R is $n \times q$ with (i, j)th entry $R_J(u_i, w_j)$, Q is a $q \times q$ matrix with (i, j)th entry $R(w_i, w_j)$, and Σ is $n \times d$ with lth column ψ_l. Letting $\check{\mathbf{d}} = (\mathbf{d}^T, \boldsymbol{\beta}^T)^T$ and $\check{S} = (S : \Sigma)$ yields the minimization of

$$(\mathbf{y} - \check{S}\check{\mathbf{d}} - R\mathbf{c})^T(\mathbf{y} - \check{S}\check{\mathbf{d}} - R\mathbf{c}) + \lambda\mathbf{c}^T Q\mathbf{c}.$$

This minimization can be performed using the method of penalized least squares described in Kim and Gu (2004).