Computational Statistics and Data Analysis 83 (2015) 236–250

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Domain selection for the varying coefficient model via local polynomial regression

Dehan Kong, Howard D. Bondell *, Yichao Wu

Department of Statistics, North Carolina State University, United States

ARTICLE INFO

Article history:
Received 20 February 2013
Received in revised form 19 March 2014
Accepted 5 October 2014
Available online 18 October 2014

Keywords:
Bandwidth selection
Oracle properties
Penalized local polynomial fitting
SCAD

ABSTRACT

In this article, we consider the varying coefficient model, which allows the relationship between the predictors and response to vary across the domain of interest, such as time. In applications, it is possible that certain predictors only affect the response in particular regions and not everywhere. This corresponds to identifying the domain where the varying coefficient is nonzero. Towards this goal, local polynomial smoothing and penalized regression are incorporated into one framework. Asymptotic properties of our penalized estimators are provided. Specifically, the estimators enjoy the oracle properties in the sense that they have the same bias and asymptotic variance as the local polynomial estimators as if the sparsity is known as *a priori*. The choice of appropriate bandwidth and computational algorithms are discussed. The proposed method is examined via simulations and a real data example.

1. Introduction

The varying coefficient model (Cleveland et al., 1991; Hastie and Tibshirani, 1993) assumes that the covariate effect may vary depending on the value of an underlying variable, such as time. It has been used in a variety of applications, such as longitudinal data analysis, and is given by

$$Y = \mathbf{x}^{\top}\mathbf{a}(U) + \epsilon, \tag{1}$$

where the predictor vector $\mathbf{x} = (x_1, \ldots, x_p)^{\top}$ represents p features, and correspondingly, $\mathbf{a}(U) = (a_1(U), \ldots, a_p(U))^{\top}$ denotes the effect of different features over the domain of the variable U. Y is the response we are interested in and ϵ denotes the random error satisfying $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2(U)$.

The varying coefficient model has been extensively studied. Many methods have been proposed to estimate its parameters. The first group of estimation methods are based on local polynomial smoothing. Examples include, but are not limited to, Fan and Gijbels (1996), Wu et al. (1998), Hoover et al. (1998), Kauermann and Tutz (1999) and Fan and Zhang (2008). The second is polynomial splines-based methods, such as Huang et al. (2002, 2004), Huang and Shen (2004) and references therein. The last group is based on smoothing splines as introduced by Hastie and Tibshirani (1993), Hoover et al. (1998), Chiang et al. (2001) and many others. In this paper, we not only consider estimation for the varying coefficient model, but also wish to identify the regions in the domain of U where predictors have an effect and the regions where they may not. This is similar, although different than variable selection, as selection methods attempt to decide whether a variable is active or not while our interest focuses on identifying regions.

* Correspondence to: 2311 Stinson Drive, Campus Box 8203, Raleigh, NC 27695-8203, United States.
E-mail address: bondell@stat.ncsu.edu (H.D. Bondell).

http://dx.doi.org/10.1016/j.csda.2014.10.004
0167-9473/

For variable selection in a traditional linear model, various shrinkage methods have been developed. They include least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), Smoothly Clipped Absolute Deviation (SCAD) (Fan and Li, 2001), adaptive LASSO (Zou, 2006) and excessively others. Although the LASSO penalty gives sparse solutions, it leads to biased estimates for large coefficients due to the linearity of the L1 penalty. To remedy this bias issue, Fan and Li (2001) proposed the SCAD penalty and showed that the SCAD penalized estimator enjoys the oracle property in the sense that not only it can select the correct submodel consistently, but also the asymptotic covariance matrix of the estimator is the same as that of the ordinary least squares estimate as if the true subset model is known as *a priori*. To achieve the goal of variable selection for group variables, Yuan and Lin (2006) developed the group LASSO penalty which penalized coefficients as a group in situations such as a factor in analysis of variance. As with the LASSO, the group LASSO estimators do not enjoy the oracle property. As a remedy, Wang et al. (2007) proposed the group SCAD penalty, which again selects variables in a group manner.

For the varying coefficient model, some existing works focus on identifying the nonzero coefficient functions, which achieves component selection for the varying coefficient functions. However, each estimated coefficient function is either zero everywhere or nonzero everywhere. For example, Wang et al. (2008) considered the varying coefficient model under the framework of a *B*-spline basis and used the group SCAD to select the significant coefficient functions. Wang and Xia (2009) combined local constant regression and the group SCAD penalization together to select the components, while Leng (2009) directly applied the component selection and smoothing operator (Lin and Zhang, 2006).

In this paper, we consider a different problem: detecting the nonzero regions for each component of the varying coefficient functions. Specifically, we aim to estimate the nonzero domain of each $a_j(U)$, which corresponds to the regions where the jth component of $\mathbf{x}$ has an effect on Y. To this end, we incorporate local polynomial smoothing together with penalized regression. More specifically, we combine local linear smoothing and group SCAD shrinkage method into one framework, which estimates not only the function coefficients but also their nonzero regions. The proposed method involves two tuning parameters, namely the bandwidth used in local polynomial smoothing and the shrinkage parameter used in the regularization method. We propose methods to select these two tuning parameters. Our theoretical results show that the resulting estimators have the same asymptotic bias and variance as the original local polynomial regression estimators.

The rest of paper is organized as follows. Section 2 reviews the local polynomial estimation for the varying coefficient model. Section 3 describes our methodology including the penalized estimation method and tuning procedure. Asymptotic properties are presented in Section 4. Simulation examples in Section 5 are used to evaluate the finite-sample performance of the proposed method. In Section 6, we apply our methods to the real data. We conclude with some discussions in Section 7.

2. Local polynomial regression for the varying coefficient model

Suppose we have independent and identically distributed (i.i.d.) samples $\{(U_i, \mathbf{x}_i^\top, Y_i)^\top, i = 1, \ldots, n\}$ from the population $(U, \mathbf{x}^\top, Y)^\top$ satisfying model (1). As $\mathbf{a}(u)$ is a vector of unspecified functions, a smoothing method must be incorporated for its estimation. In this article, we adopt the local linear smoothing for this varying coefficient model (Fan and Zhang, 1999). For U in a small neighborhood of u, we can approximate the function $a_j(U)$, $1 \le j \le p$, by a linear function

$$a_j(U) \approx a_j(u) + a_j'(u)(U - u).$$

For a fixed point u, denote $a_j(u)$ and $a_j'(u)$ by a_j and b_j, respectively, and denote their estimates by $\hat{a}_j$ and $\hat{b}_j$, which estimate the function $a_j(\cdot)$ and its derivative at the point u. Note that $(\hat{a}_j, \hat{b}_j)$ $(1 \le j \le p)$ can be estimated via local polynomial regression by solving the following optimization problem:

$$\min_{\mathbf{a},\mathbf{b}} \sum_{i=1}^{n} \{Y_i - \mathbf{x}_i^\top \mathbf{a} - \mathbf{x}_i^\top \mathbf{b}(U_i - u)\}^2 (K_h(U_i - u)/K_h(0)), \quad (2)$$

where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$, $K_h(t) = K(t/h)/h$, and $K(t)$ is a kernel function. The parameter $h > 0$ is the bandwidth controlling the size of the local neighborhood. It implicitly controls the model complexity. Consequently it is essential to choose an appropriate smoothing bandwidth in local polynomial regression. We will discuss how to select the bandwidth h in Section 2.1.

The kernel function $K(\cdot)$ is a nonnegative symmetric density function satisfying $\int K(t)dt = 1$. There are numerous choices for the kernel function. Examples are Gaussian kernel ($K(t) = \exp(-t^2/2)/\sqrt{2\pi}$) and Epanechnikov kernel ($K(t) = 0.75(1 - t^2)_+$) among many others. Typically, the estimates are not sensitive to the choice of the kernel function. In this paper, we use the Epanechnikov kernel, which leads to computational efficiency due to its bounded support.

Notice here that our loss function is slightly different from the loss function of the traditional local polynomial regression for the varying coefficient model (Fan and Zhang, 1999). We have rescaled the original loss function by a term $K_h(0)$. For a fixed h, this change does not affect the estimates. However, this scaling is needed later to correctly balance the loss function and penalty term since $K_h(U_i - u) = K((U_i - h)/h)/h$, we include the term $K_h(0)$ to eliminate the effect of h so that $K_h(U_i - u)/K_h(0) = O(1)$.

Denote $\mathbf{a}_0 = (a_{01}, \ldots, a_{0p})^\top$ and $\mathbf{b}_0 = (b_{01}, \ldots, b_{0p})^\top$ to be the true values of the coefficient functions and their derivatives, respectively, and $\hat{\mathbf{a}} = (\hat{a}_{01}, \ldots, \hat{a}_{0p})^\top$ and $\hat{\mathbf{b}} = (\hat{b}_{01}, \ldots, \hat{b}_{0p})^\top$ as their corresponding local polynomial regression