Statistics and Probability Letters 83 (2013) 2324–2331

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Shrinkage estimation of partially linear single-index models

Qin Wang[a,*], Rongning Wu[b]

[a] Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA 23284, USA

[b] Department of Statistics and Computer Information Systems, Baruch College, The City University of New York, New York, NY 10010, USA

ARTICLE INFO

Article history:
Received 4 February 2013
Received in revised form 12 February 2013
Accepted 15 June 2013
Available online 22 June 2013

Keywords:
Adaptive Lasso
Partially linear model
Semiparametric estimation

ABSTRACT

In this paper, we propose shrinkage estimation for partially linear single-index models. A profile least squares approximation is used to estimate the model parameters and select informative variables simultaneously. The resulting estimator is shown to be consistent and to enjoy the oracle properties.

1. Introduction

Partially linear single-index models (PLSIMs) combine naturally the advantages of classical linear models and those of nonparametric regression models. Due to its flexibility and generality, this class of models has gained much attention since its introduction by Carroll et al. (1997). Let Y be the response variable, and let $\mathbf{Z}$ and $\mathbf{X}$ be two covariate vectors of length p and q, respectively. Following Xia and Härdle (2006), we formulate a PLSIM as

$$Y = \eta(\boldsymbol{\alpha}^T\mathbf{Z}) + \boldsymbol{\beta}^T\mathbf{X} + \epsilon, \tag{1}$$

where ϵ is the error term such that $E(\epsilon \mid \mathbf{Z}, \mathbf{X}) = 0$ and $\mathrm{Var}(\epsilon \mid \mathbf{Z}, \mathbf{X}) < \infty$ almost surely, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^T$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T$ are parameter vectors, and $\eta(\cdot)$ is an unknown smooth function. For identifiability, it is assumed that $\|\boldsymbol{\alpha}\| = 1$ and that the first entry of $\boldsymbol{\alpha}$ is positive. The use of the single-index $\boldsymbol{\alpha}^T\mathbf{Z}$ is to circumvent the "curse of dimensionality" in multivariate nonparametric regression analysis.

For PLSIMs, several estimation approaches have been proposed in the literature in addition to the backfitting algorithm of Carroll et al. (1997). For example, Yu and Ruppert (2002) proposed a penalized spline procedure, Xia and Härdle (2006) studied the minimum average variance estimation (MAVE) method, Wang et al. (2010) proposed a two-stage procedure under the extra condition that $\mathbf{X}$ can be explained by a subset of indices based on $\mathbf{Z}$, and Liang et al. (2010) employed a profile least squares (PrLS) approach.

In practice, a parsimonious model that retains only significant covariates is more desirable. The lasso technique, initially developed by Tibshirani (1996) for linear regression models, selects the significant variables and estimates the regression coefficients simultaneously. Employing an L_1-type penalty function on the regression coefficients to shrink their estimates, it can produce the correct model asymptotically when the model is sparse, a property called consistency in variable selection. Recently, Zou (2006) proposed an adaptive lasso procedure, in which insignificant variables are penalized heavily whereas significant variables are penalized relatively slightly, and showed that the adaptive procedure further enjoys the oracle

* Corresponding author. Tel.: +1 8048281316.
E-mail addresses: qwang3@vcu.edu (Q. Wang), rongning.wu@baruch.cuny.edu (R. Wu).

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.06.019

properties. That is, the resulting estimator attains the full efficiency of the estimator obtained as if the true model were known in advance. This backs the popularity of lasso-type methods due to their excellent empirical performance.

In this paper, we apply the adaptive lasso to PLSIMs. A shrinkage procedure based on a profile least squares approximation (PrLSA) is proposed to simultaneously estimate the model parameters and select informative variables in both parametric and nonparametric components. Since the procedure essentially solves a convex optimization problem, it is computationally more attractive while being as efficient as the aforementioned methods. On the other hand, we show that the resulting estimator is consistent in both parameter estimation and variable selection, and also that it enjoys the oracle properties. Asymptotically, the zero regression coefficients are estimated to be zero almost surely, and the estimator of nonzero regression coefficients is normal and performs as if the true model were known. Therefore, the proposed procedure is a valuable supplement to existing methods.

The remainder of the paper is organized as follows. Section 2 develops the implementation algorithm of the PrLSA procedure and its penalized version. The asymptotic properties of the resulting estimator are also presented, with technical proofs given in the Appendix. In Section 3, simulation studies are conducted to evaluate the finite-sample performance of the procedure.

2. Profile least squares based shrinkage estimation

2.1. Method and algorithm

Let $\{(Y_i, \mathbf{Z}_i, \mathbf{X}_i), i = 1, \ldots, n\}$ be an independent and identically distributed (i.i.d.) random sample from the model (1) with true parameter values $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and link function $\eta_0(\cdot)$, namely $Y_i = \eta_0(\boldsymbol{\alpha}_0^T\mathbf{Z}_i)+\boldsymbol{\beta}_0^T\mathbf{X}_i+\epsilon_i$. This is a typical semiparametric setting as $\eta_0(\cdot)$ is modeled nonparametrically.

We propose a PrLSA procedure to estimate $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\eta(\cdot)$. First of all, at each data point $(\mathbf{z}, \mathbf{x})$, one can estimate $\eta(\boldsymbol{\alpha}^T\mathbf{z})$ using local linear regression (Fan and Gijbels, 1996) by minimizing

$$\sum_{i=1}^{n}\left[Y_i - \boldsymbol{\beta}^T\mathbf{X}_i - a - b(\boldsymbol{\alpha}^T(\mathbf{Z}_i - \mathbf{z}))\right]^2 K_h(\boldsymbol{\alpha}^T(\mathbf{Z}_i - \mathbf{z}))$$

with respect to (a, b), where $K_h(\cdot) = K(\cdot/h)/h$, with $K(\cdot)$ being a symmetric kernel function and h being the bandwidth. A closed-form solution can be obtained as

$$\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} = \left\{\sum_{i=1}^{n} K_h(\boldsymbol{\alpha}^T(\mathbf{Z}_i - \mathbf{z}))\mathcal{Z}_i\mathcal{Z}_i^T\right\}^{-1}\left\{\sum_{i=1}^{n}(Y_i - \boldsymbol{\beta}^T\mathbf{X}_i)K_h(\boldsymbol{\alpha}^T(\mathbf{Z}_i - \mathbf{z}))\mathcal{Z}_i\right\}, \tag{2}$$

where $\mathcal{Z}_i = (1, \boldsymbol{\alpha}^T(\mathbf{Z}_i - \mathbf{z}))^T$. $\eta(\boldsymbol{\alpha}^T\mathbf{z})$ is estimated by $\hat{a} = \hat{\eta}(\boldsymbol{\alpha}^T\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\beta})$, a function of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Moreover, partial derivatives of $\hat{\eta}(\boldsymbol{\alpha}^T\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are readily derived; they are also functions of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

Now, to estimate $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, we formulate an ordinary least squares (OLS)-type objective function rather than minimizing

$$Q(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{j=1}^{n}\left[Y_j - \boldsymbol{\beta}^T\mathbf{X}_j - \hat{\eta}(\boldsymbol{\alpha}^T\mathbf{Z}_j; \boldsymbol{\alpha}, \boldsymbol{\beta})\right]^2. \tag{3}$$

To be specific, a local linear approximation of $\hat{\eta}(\boldsymbol{\alpha}^T\mathbf{Z}_j; \boldsymbol{\alpha}, \boldsymbol{\beta})$ around an initial value $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$ yields

$$\hat{\eta}(\boldsymbol{\alpha}^T\mathbf{Z}_j; \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx \hat{\eta}(\tilde{\boldsymbol{\alpha}}^T\mathbf{Z}_j; \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}) + \left.\frac{\partial\hat{\eta}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)}\right|_{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})}\begin{pmatrix}\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}\\ \boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}\end{pmatrix},$$

where

$$\left.\frac{\partial\hat{\eta}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)}\right|_{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})} := \left.\left(\frac{\partial\hat{\eta}(\boldsymbol{\alpha}^T\mathbf{Z}_j; \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial\boldsymbol{\alpha}^T}, \frac{\partial\hat{\eta}(\boldsymbol{\alpha}^T\mathbf{Z}_j; \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\right)\right|_{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})}.$$

Then, $Q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be approximated by

$$Q_L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{j=1}^{n}\left[Y_j - \boldsymbol{\beta}^T\mathbf{X}_j - \hat{\eta}(\tilde{\boldsymbol{\alpha}}^T\mathbf{Z}_j; \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}) - \left.\frac{\partial\hat{\eta}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)}\right|_{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})}\begin{pmatrix}\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}\\ \boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}\end{pmatrix}\right]^2.$$

Letting

$$Y_j^* = Y_j - \hat{\eta}(\tilde{\boldsymbol{\alpha}}^T\mathbf{Z}_j; \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}) + \left.\frac{\partial\hat{\eta}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)}\right|_{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})}\begin{pmatrix}\tilde{\boldsymbol{\alpha}}\\ \tilde{\boldsymbol{\beta}}\end{pmatrix},$$

we have

$$Q_L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{j=1}^{n}\left[Y_j^* - \left(\left.\frac{\partial\hat{\eta}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)}\right|_{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})} + (\mathbf{0}_{1\times p}, \mathbf{X}_j^T)\right)\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{pmatrix}\right]^2. \tag{4}$$

Minimizing (4) with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, we obtain an exact OLS closed-form solution.