Journal of the Korean Statistical Society 44 (2015) 95–110

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Walsh-average based variable selection for varying coefficient models

Kangning Wang[a,b,c,*], Lu Lin[b,c]

[a] Department of Mathematics & KLDAIP, Chongqing University of Arts and Sciences, Chongqing, China
[b] Shandong University Qilu Securities Institute for Financial Studies, Shandong University, Jinan, China
[c] School of Mathematics, Shandong University, Jinan, China

ARTICLE INFO

Article history:
Received 18 January 2014
Accepted 26 May 2014
Available online 18 June 2014

AMS 2000 subject classifications:
primary 62G05
secondary 62E20

Keywords:
Varying coefficient model
Walsh-average
Oracle property
Variable selection
Asymptotic relative efficiency
Robust BIC-type criterion

ABSTRACT

A shrinkage-type variable selection procedure for varying coefficient models is routinely established in the least-squares (LS) framework. Although the LS method has favorable properties for a large class of error distributions, it will break down if the error variance is infinite and is adversely affected by outliers and heavy-tail distributions. To overcome these issues, we propose a robust shrinkage method termed regularized Walsh-average (RWA) that can construct robust nonparametric variable selection and robust coefficient estimation simultaneously. Theoretical analysis reveals RWA works beautifully, including consistency in variable selection and oracle property in estimation, even when error variance is infinite. More important property is that when error variance is finite, compared with the LS based estimators, the asymptotic relative efficiency of the new estimator is at least 0.8896, a relatively high level. Furthermore, a robust BIC-type criterion, which can identify the true model consistently, is suggested for shrinkage parameter selection. Numerical studies also confirm our theories.

1. Introduction

1.1. Background

Suppose a random sample $\{(U_i, \mathbf{X}_i^\top, Y_i), i = 1, \ldots, n\}$ comes from the following varying coefficient (VC) model:

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta}(U_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_p(\cdot))^\top$ whose true value is $\boldsymbol{\beta}_0(\cdot) = (\beta_{01}(\cdot), \ldots, \beta_{0p}(\cdot))^\top$, $Y_i \in \mathbb{R}$ is the response variable, $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^\top \in \mathbb{R}^p$ is a p-dimensional covariate and $U_i \in [0, 1]$ is an index variable with density function $f_U(u)$, $\epsilon_i \in \mathbb{R}$ is the random error following some pdf $f(\cdot)$. Let $\mathcal{A} = \{j : 0 < E\{\beta_{0j}^2(U_i)\} < \infty, j = 1, \ldots, p\}$ and $\bar{\mathcal{A}} = \{j : E\{\beta_{0j}^2(U_i)\} = 0, j = 1, \ldots, p\}$. Simply speaking, the covariate $\boldsymbol{X}_{i\mathcal{A}} = (X_{ij}, j \in \mathcal{A})^\top \in \mathbb{R}^{|\mathcal{A}|}$ is truly relevant but the rest are not, where $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$. Let $\boldsymbol{\beta}_{\mathcal{A}}(\cdot)$ and $\boldsymbol{\beta}_{0\mathcal{A}}(\cdot)$ be the subvectors of $\boldsymbol{\beta}(\cdot)$ and $\boldsymbol{\beta}_0(\cdot)$ respectively, whose entries correspond to the relevant variable in $\mathcal{A}$. The goal of our work is to select the relevant variables and estimate the unknown nonzero coefficients efficiently and robustly.

* Corresponding author at: Department of Mathematics & KLDAIP, Chongqing University of Arts and Sciences, Chongqing, China.
E-mail address: wkn1986@126.com (K. Wang).

http://dx.doi.org/10.1016/j.jkss.2014.05.006
1226-3192/

Shrinkage-type methods have recently seen increasing applications in variable selection. In linear regression, which include but not limited to the nonnegative garrotte (Breiman, 1995; Yuan & Lin, 2007), the LASSO (Tibshirani, 1996; Wang, Li, & Tsai, 2007; Zou, 2006), the bridge regression (Fu, 1998; Knight & Fu, 2000), and the SCAD (Fan & Li, 2001). Furthermore, robust shrinkage methods for linear regression also have been well studied (e.g., Leng, 2010; Wang & Li, 2009; Wang et al., 2007; Zou & Yuan, 2008). Recent years, a number of works have been done to extend the shrinkage estimation methods to nonparametric and semiparametric settings. Fan and Li (2004) extended the SCAD to partially linear models. Li and Liang (2008) studied variable selection for partially linear VC models via the SCAD and the generalized likelihood test. Wang, Li, and Huang (2008) and Wang and Xia (2009) proposed two shrinkage methods for the VC model, which combine the least-squares (LS) and penalty to estimate the coefficient functions and select important nonparametric components automatically. Tang, Wang, Zhu, and Song (2012) developed a variable selection approach for both LS regression and quantile regression models with possibly varying coefficients via basis expansion and an adaptive-LASSO-type penalty. Kai, Li, and Zou (2011) proposed a penalized semiparametric composite quantile regression (CQR) method to select variables only for parameter parts in partially linear VC models.

1.2. Issues

However, shrinkage variable selection methods for the VC model are mainly built on the LS method. Although the LS based method is successful and remains a good asymptotical property for a large class of random error distributions, its defects cannot be ignored. Firstly, the finite variance assumption is crucial; without this assumption (e.g., the error follows Cauchy distribution), the LS estimator is no longer consistent. The second and more important issue is that even the finite variance assumption is satisfied, their efficiency can deteriorate dramatically when the error is subject to heavy-tailed distribution. Thirdly, these estimators are also very sensitive to outliers in the data.

These considerations motivate us to develop a robust nonparametric shrinkage method, which can remedy the defects of the LS based method, and given the popularity of the kernel smoothing, it is desirable to have such a method that can work with kernel smoothing techniques in a very natural way. But it is not a trivial task. At first glance, a natural alternative seems to be the quantile together with a penalty (e.g. Kai et al., 2011; Noh, Chung, & Keilegom, 2012; Tang, Wang, & Zhu, 2013; Zhao, Zhang, Lv, & Liu, 2012), but as shown by Feng, Zou, and Wang (2012), it has limitations in terms of efficiency and uniqueness of estimation. Furthermore, how to design a shrinkage parameter selection criterion that not only has beautiful theoretical property but also can be robust, is also a interesting issue that has been neglected in the aforementioned literatures.

1.3. Our contributions

In this paper, we propose an efficient and robust nonparametric shrinkage method for the VC model (1.1), referred to as the regularized Walsh-average (RWA) estimator. Compared with the LS based method, the new RWA has some desirable features.

(1) According to normal approximation and penalized approximate likelihood, the RWA estimator uses the covariance information of the initial estimator sufficiently, this makes the new method more efficient and robust, and this is the main innovation compared with the traditional method.
(2) Even when error variance is infinite, the new RWA method works beautifully, including consistency in variable selection and oracle property in estimation.
(3) When the error variance is finite, the new approach can substantially improve upon the LS based procedure for a wide class of error distribution. Theoretical analysis reveals that in comparison with the LS estimators, the asymptotic relative efficiency (ARE) of the new RWA method, measured by the asymptotic mean squared error of the nonzero coefficient functions estimator, has a lower bound 0.8896.

Furtherly, we propose a robust BIC-type criterion, a robust fitting error criterion, to choose the penalty parameter. We theoretically prove that, based on the shrinkage parameter selected by the proposed BIC-type criterion, the new RWA can identify the true model consistently.

The rest of this paper is organized in the following way. In Section 2, the RWA method and its algorithm are presented. The theoretical properties are carefully studied in Section 3. In Section 4, we propose a robust BIC-type information criterion. Simulation studies are given in Section 5. All technical details are postponed to the Appendix.

2. Methodology and algorithm

2.1. Regularized Walsh-average estimation

Let $\boldsymbol{\beta}_t = \boldsymbol{\beta}(U_t)$ with true value $\boldsymbol{\beta}_{0t} = \boldsymbol{\beta}_0(U_t)$, $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n)^\top = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p) \in \mathbb{R}^{n\times p}$ with true value $\boldsymbol{B}_0 = (\boldsymbol{\beta}_{01}, \ldots, \boldsymbol{\beta}_{0n})^\top = (\boldsymbol{b}_{01}, \ldots, \boldsymbol{b}_{0p})$, where $\boldsymbol{b}_j = (\beta_j(U_1), \ldots, \beta_j(U_n))^\top \in \mathbb{R}^n$ with true value $\boldsymbol{b}_{0j} = (\beta_{0j}(U_1), \ldots, \beta_{0j}(U_n))^\top \in \mathbb{R}^n$. It is obvious by the definition of $\bar{\mathcal{A}}$ that $\|\boldsymbol{b}_{0j}\|_2 = 0, j \in \bar{\mathcal{A}}$, where $\|\cdot\|_2$ is the usual Euclidean norm. Thus, selecting the relevant